ADVANCES IN MATHEMATICS 29, 178–218 (1978)

The Diamond Lemma for Ring Theory*

GEORGE M. BERGMAN

Department of Mathematics, University of California, Berkeley, California 94720

Contents. Introduction. 1. The main theorem: Bases for associative k-algebras. 2. Application: A problem on idempotents. 3. Application: The Poincaré–Birkhoff–Witt theorem. 4. Applications: How to take advantage of identities. 5. Observations on some practical questions. 6. Generalization: Bimodule-structures of k-rings. 7. Generalization: Right bases for k-rings. 8. Applications: Coproducts of k-rings. 9. Analogs: Semigroups, modules, additive categories, etc. 10. Contrasts: Other sorts of algebraic structures. 11. General observations on canonical form results.

INTRODUCTION

The main results in this paper are trivial. But what is trivial when described in the abstract can be far from clear in the context of a complicated situation where it is needed. Hence it seems worthwhile to set down explicit formulations and proofs of these results.

We will be concerned with the conditions that must be verified to establish a canonical form for the elements of a ring, semigroup, or similar algebraic structure.

Suppose R is an associative algebra with 1 over the commutative ring k, and that we have a presentation of R by a family X of generators and a family S of relations. Suppose each relation $\sigma \in S$ has been written in the form $W_\sigma = f_\sigma$, where W_σ is a monomial (a product of elements of X) and f_σ is a k-linear combination of monomials, and that we wish to use these relations as instructions for "reducing" (or "straightening") expressions r for elements of R. That is, if any of the monomials occurring in the expression r contains one of the W_σ as a subword, we substitute f_σ for that subword, and we iterate this process as long as possible. Now the difficulty is that this process is not in general well defined. At each step we must choose which reduction to apply to which subword of which monomial. Under what hypotheses can one nonetheless show that such a procedure will bring every expression to a unique irreducible form? And may one then conclude that these yield a canonical form for elements of R?

* Part of this work was done while the author held a National Science Foundation Graduate Fellowship, part while he was partly supported by NSF contract GP 9152, and part while he was supported by NSF contract MPS 73-08528.

0001-8708/78/0292-0178$05.00/0

There is a very general result of this sort due to Newman [43, Sect. 3], often called the "Diamond Lemma." Let G be an oriented graph. Here the vertices of G may be expressions for the elements of some algebraic object and the edges reduction steps going from one such expression to another "better" one, where enough reductions are given so that the equivalence relation they generate corresponds to equality in the object. Now suppose that

(i) The oriented graph G has descending chain condition. That is, all positively oriented paths in G terminate; and

(ii) whenever two edges, e and e', proceed from one vertex a of G, there exist positively oriented paths p, p' in G leading from the endpoints b, b' of these edges to a common vertex c. (The "confluence" or "diamond" condition.)

Then every connected component C of G has a unique minimal vertex m_C. This means that every maximal positively oriented path beginning at a point of C will terminate at m_C; in other words, that the given reduction procedure yields unique canonical forms for elements of the original algebraic object.

Our main result, Theorem 1.2, is an analog and strengthening of the above observations for the case of associative rings, with reduction procedures of the form sketched earlier. It is self-contained and does not follow Newman's graph-theoretic formulation. The strengthening lies in the result that the analogs of conditions (i) and (ii) need only be verified for *monomials*, and in fact that (ii) need only be verified for "*minimal* nontrivial ambiguously reducible monomials". That is, it suffices to check, for each monomial which can be written as ABC with either $AB = W_\sigma$, $BC = W_\tau$ ($\sigma, \tau \in S, B \neq 1$) or $ABC = W_\sigma, B = W_\tau$ ($\sigma \neq \tau \in S$) that the two expressions to which ABC reduces (in the first case, $f_\sigma C$ and Af_τ; in the second f_σ and $Af_\tau C$) can be reduced to a common value.

This fact has been considered obvious and used freely by some ring-theorists (e.g., [17, Sect. 5]), but others seem unaware of it and write out tortuous verifications. Our proof is straightforward, but not quite as vacuous as it might seem, owing to the complication that after a substitution in a polynomial expression, some of the newly arising terms can coincide with monomials already occurring.

A further improvement we get, useful in many applications, is that in the analog of (ii), the paths p and p' need not be strictly positively oriented, but must merely stay "below" the original element ABC, with respect to a partial ordering on monomials introduced in connection with (i). (Condition (a') of Theorem 1.2.) In Section 4 we also show how one may take advantage of identities to greatly simplify the verification of the confluence conditions in certain common sorts of presentation. Other sections give applications, variants, and generalizations of the main result.

Sections 1, 2.1, 3, (4) form a natural unit for the reader or seminar that does not want to take on this long a paper; a sort of "What every ring theorist should know." The reader who wants to go on from there but skip the more specialized

and tedious parts might then read Sections 5, 6, the first half of 8, 9.1–9.5, 10.1, 10.3, 11.2, and 11.3, with modifications depending on his or her own taste and interest. The semigroup-theorist might want to read only Sections 1, 5, 9.1, 9.2, and 11.2. All sections depend on Section 1; subsequent sections are largely independent of one another, except for the sequences 3–4 and 6–8.

I am indebted to P. M. Cohn and others for acquainting me with relevant literature.

1. The Main Theorem: Bases for Associative k-Algebras

Let k be a commutative associative ring with 1, X a set, $\langle X\rangle$ the free semigroup with 1 on X, and $k\langle X\rangle$ the free associative k-algebra on X, which is the semigroup algebra of $\langle X\rangle$.

Let S be a set of pairs of the form $\sigma = (W_\sigma, f_\sigma)$, where $W_\sigma \in \langle X\rangle, f_\sigma \in k\langle X\rangle$. For any $\sigma \in S$ and $A, B \in \langle X\rangle$, let $r_{A\sigma B}$ denote the *k-module* endomorphism of $k\langle X\rangle$ that fixes all elements of $\langle X\rangle$ other than $AW_\sigma B$, and that sends this basis element to $Af_\sigma B$. We shall call the given set S a *reduction system*, and the maps $r_{A\sigma B}\colon k\langle X\rangle \to k\langle X\rangle$ *reductions*.

We shall say a reduction $r_{A\sigma B}$ acts *trivially* on an element $a \in k\langle X\rangle$ if the coefficient of $AW_\sigma B$ in a is zero, and we shall call a *irreducible* (under S) if every reduction is trivial on a, i.e., if a involves none of the monomials $AW_\sigma B$. The k-submodule of all irreducible elements of $k\langle X\rangle$ will be denoted $k\langle X\rangle_{\mathrm{irr}}$. A finite sequence of reductions $r_1, \ldots, r_n$ ($r_i = r_{A_i\sigma_i B_i}$) will be said to be *final* on $a \in k\langle X\rangle$ if $r_n \cdots r_1(a) \in k\langle X\rangle_{\mathrm{irr}}$.

An element a of $k\langle X\rangle$ will be called *reduction-finite* if for every infinite sequence $r_1, r_2, \ldots$ of reductions, r_i acts trivially on $r_{i-1} \cdots r_1(a)$ for all sufficiently large i. If a is reduction-finite, then any maximal sequence of reductions r_i, such that each r_i acts *nontrivially* on $r_{i-1} \cdots r_1(a)$, will be finite, and hence a final sequence. It follows from their definition that the reduction-finite elements form a k-submodule of $k\langle X\rangle$.

We shall call an element $a \in k\langle X\rangle$ *reduction-unique* if it is reduction-finite, and if its images under all final sequences of reductions are the same. This common value will be denoted $r_S(a)$.

The next lemma will handle the difficulties in the proof of our theorem that would arise from possible coalescence and cancellation of terms after reduction of a sum or product.

Lemma 1.1. (i) *The set of reduction-unique elements of $k\langle X\rangle$ forms a k-submodule, and r_S is a k-linear map of this submodule into $k\langle X\rangle_{\mathrm{irr}}$.*

(ii) *Suppose $a, b, c \in k\langle X\rangle$ are such that for all monomials A, B, C occurring with nonzero coefficient in a, b, c, respectively, the product ABC is reduction-unique.*

(In particular this implies that abc is reduction-unique.) Let r be any finite composition of reductions. Then ar(b)c is reduction-unique, and $r_S(ar(b)c) = r_S(abc)$.

Proof. (i) Say $a, b \in k\langle X\rangle$ are reduction-unique, and $\alpha \in k$. We know $\alpha a + b$ is reduction-finite. Let r be any composition of reductions final on this element. Since a is reduction-unique we can find a composition of reductions r' such that $r'r(a) = r_S(a)$, and similarly there is a composition of reductions r'' such that $r''r'r(b) = r_S(b)$. As $r(\alpha a + b)$ is irreducible, we have $r(\alpha a + b) = r''r'r(\alpha a + b) = \alpha r''r'r(a) + r''r'r(b) = \alpha r_S(a) + r_S(b)$, from which our assertions follow.

(ii) By (i) and the way (ii) is formulated, it clearly suffices to prove (ii) in the case where a, b, c are monomials A, B, C, and r is a single reduction $r_{D\sigma E}$. But in this case, $Ar_{D\sigma E}(B)C = r_{AD\sigma EC}(ABC)$, which is the image of ABC under a reduction, hence is reduction-unique if ABC is, with the same reduced form. ∎

Let us call a 5-tuple (σ, τ, A, B, C) with $\sigma, \tau \in S$ and $A, B, C \in \langle X\rangle - \{1\}$, such that $W_\sigma = AB$, $W_\tau = BC$, an *overlap ambiguity* of S. We shall say the overlap ambiguity (σ, τ, A, B, C) is *resolvable* if there exist compositions of reductions, r and r', such that $r(f_\sigma C) = r'(Af_\tau)$ (the confluence condition on the results of the two indicated ways of reducing ABC.)

Similarly, a 5-tuple (σ, τ, A, B, C) with $\sigma \neq \tau \in S$ and $A, B, C \in \langle X\rangle$ will be called an *inclusion ambiguity* if $W_\sigma = B$, $W_\tau = ABC$; and such an ambiguity will be called resolvable if $Af_\sigma C$ and f_τ can be reduced to a common expression.

By a *semigroup partial ordering* on $\langle X\rangle$ we shall mean a partial order "$\leqslant$" such that $B < B' \Rightarrow ABC < AB'C$ $(A, B, B', C \in \langle X\rangle)$, and it will be called *compatible* with S if for all $\sigma \in S, f_\sigma$ is a linear combination of monomials $< W_\sigma$.

Let I denote the two-sided ideal of $k\langle X\rangle$ generated by the elements $W_\sigma - f_\sigma$ $(\sigma \in S)$. As a k-module, I is spanned by the products $A(W_\sigma - f_\sigma)B$.

If $\leqslant$ is a partial order on $\langle X\rangle$ compatible with the reduction system S, and A is any element of $\langle X\rangle$, let I_A denote the submodule of $k\langle X\rangle$ spanned by all elements $B(W_\sigma - f_\sigma)C$ such that $BW_\sigma C < A$. We shall say that an ambiguity (σ, τ, A, B, C) is resolvable *relative to* $\leqslant$ if $f_\sigma C - Af_\tau \in I_{ABC}$ (or for inclusion ambiguities, if $Af_\sigma B - f_\tau \in I_{ABC}$). Any resolvable ambiguity is resolvable relative to $\leqslant$.

THEOREM 1.2. *Let S be a reduction system for a free associative algebra $k\langle X\rangle$ (a subset of $\langle X\rangle \times k\langle X\rangle$), and $\leqslant$ a semigroup partial ordering on $\langle X\rangle$, compatible with S, and having descending chain condition. Then the following conditions are equivalent:*

(a) *All ambiguities of S are resolvable.*

(a′) *All ambiguities of S are resolvable relative to $\leqslant$.*

(b) *All elements of $k\langle X\rangle$ are reduction-unique under S.*

(c) *A set of representatives in $k\langle X\rangle$ for the elements of the algebra $R = k\langle X\rangle/I$ determined by the generators X and the relations $W_\sigma = f_\sigma$ ($\sigma \in S$) is given by the k-submodule $k\langle X\rangle_{\mathrm{irr}}$ spanned by the S-irreducible monomials of $\langle X\rangle$.*

When these conditions hold, R may be identified with the k-module $k\langle X\rangle_{\mathrm{irr}}$, made a k-algebra by the multiplication $a \cdot b = r_S(ab)$.

Proof. We easily see from our general hypothesis, by induction with respect to the partial ordering with descending chain condition $\leqslant$, that every element of $\langle X\rangle$, and hence every element of $k\langle X\rangle$, is reduction-finite.

To show (b) $\Leftrightarrow$ (c), first note that (c) simply says $k\langle X\rangle = k\langle X\rangle_{\mathrm{irr}} \oplus I$. Assuming (b), r_S will be a projection of $k\langle X\rangle$ onto $k\langle X\rangle_{\mathrm{irr}}$; its kernel is contained in I because every reduction alters an element by a member of I, and contains I because for all A, B, σ, $r_S(A(W_\sigma - f_\sigma)B) = r_S(AW_\sigma B) - r_S(Af_\sigma B) = 0$ by Lemma 1.1, proving (c). Conversely, assume (c) and suppose $a \in k\langle X\rangle$ can be reduced to either of b, $b' \in k\langle X\rangle_{\mathrm{irr}}$. Then $b - b' \in k\langle X\rangle_{\mathrm{irr}} \cap I = \{0\}$, proving (b).

The final comment in the statement of the theorem is clear, and the implications (b) $\Rightarrow$ (a) $\Rightarrow$ (a') are immediate. It remains to prove (a') $\Rightarrow$ (b).

Assume (a'). It will suffice to prove all monomials $D \in \langle X\rangle$ reduction-unique, since the reduction-unique elements of $k\langle X\rangle$ form a submodule. We assume inductively that all monomials $< D$ are reduction-unique. Thus the domain of r_S includes the submodule spanned by all these monomials, so the kernel of r_S contains I_D. We must now show that given any two reductions $r_{L\sigma M'}$ and $r_{L'\tau M}$ each acting nontrivially on D (and hence each sending D to a linear combination of monomials $< D$) we will have $r_S(r_{L\sigma M'}(D)) = r_S(r_{L'\tau M}(D))$. There are three cases, according to the relative locations of the subwords W_σ and W_τ in the monomial D. We may assume without loss of generality that length$(L) \leqslant$ length (L'), i.e., that the indicated copy of W_σ in D begins no later than the indicated copy of W_τ.

Case 1. The subwords W_σ and W_τ overlap in D, but neither contains the other. Then $D = LABCM$, where (σ, τ, A, B, C) is an overlap ambiguity of S, and $r_{L\sigma M'}(D) - r_{L'\tau M}(D) = L(f_\sigma C - Af_\tau)M$. By (a'), $f_\sigma C - Af_\tau \in I_{ABC}$, so $L(f_\sigma C - Af_\tau)M \in I_{LABCM} = I_D$, which is annihilated by r_S. Thus $r_S(r_{L\sigma M'}(D)) - r_S(r_{L'\tau M}(D)) = 0$, as required.

Case 2. One of the subwords W_σ, W_τ of D is contained in the other. This case is handled like the preceding, using the resolvability of *inclusion* ambiguities relative to $\leqslant$.

Case 3. W_σ and W_τ are disjoint subwords of D. Then $D = LW_\sigma NW_\tau M$, and the elements we must prove equal are $r_S(Lf_\sigma NW_\tau M)$ and $r_S(LW_\sigma Nf_\tau M)$. But Lemma 1.1(ii) shows each of these equal to $r_S(Lf_\sigma Nf_\tau M)$. ∎

(One might ask: If a reduction system S has unresolvable ambiguities, can't we *make* the reduction-procedure well-defined by restricting our reduction procedure by some additional rules which specify which reduction to apply whenever there is a choice? This *would* give a well-defined procedure, but not a canonical form. Two irreducible expressions to which one element could have been reduced under the unrestricted procedure will still be distinct irreducible expressions corresponding to the same element of R. We shall see in the next section that the way to handle unresolvable ambiguities is not to restrict but to extend our reduction procedure.)

The machinery set up above is of course just a formalization of very natural considerations in the study of normal forms for elements of a k-algebra. Having proved things formally, we shall in our applications often return to informal language, e.g., speak of substituting a certain expression for a certain word rather than saying that a certain pair will belong to our reduction system.

In this paper we are mainly interested in the canonical forms themselves, but the following obvious Corollary may be of use for other purposes:

COROLLARY 1.3. *Let $k\langle X\rangle$ be a free associative algebras, and "$\leqslant$" a semigroup partial ordering of $\langle X\rangle$ with descending chain condition.*

If S is a reduction system on $k\langle X\rangle$ compatible with $\leqslant$ and having no *ambiguities, then the set of k-algebra relations $W_\sigma = f_\sigma$ $(\sigma \in S)$ is* independent.

More generally, if $S_1 \subseteq S_2$ are reduction systems, such that S_1 is compatible with $\leqslant$ and all its ambiguities are resolvable, and if S_2 contains some σ such that W_σ is irreducible with respect to S_1, then the inclusion of ideals associated with these systems, $I_1 \subseteq I_2$, is strict. ∎

2. APPLICATION: A PROBLEM ON IDEMPOTENTS

2.1. I was first led to the ideas of this paper by struggling with:

American Mathematical Monthly ADVANCED PROBLEM 5082 [1]. *Let R be a ring in which, if either $x + x = 0$ or $x + x + x = 0$, it follows that $x = 0$. Suppose that a, b, c and $a + b + c$ are all idempotents in R. Does it follow that $ab = 0$?*

Evidently we should study the ring R defined by generators a, b, c and relations:

$$a^2 = a, \tag{1}$$

$$b^2 = b, \tag{2}$$

$$c^2 = c, \tag{3}$$

$$(a + b + c)^2 = a + b + c. \tag{4$_0$}$$

We can use the first three relations to eliminate monomials containing the sequences *aa*, *bb*, *cc* from all expressions in R. If in (4_0) we expand the left-hand side, simplify the result using (1), (2) and (3), and isolate an arbitrary term, we get the relation:

$$ba = -ab - bc - cb - ac - ca. \tag{4}$$

Do (1)–(4), used as reduction formulas, yield unique canonical forms for elements of R? There are five ambiguously reducible sequences:

$$aaa, bbb, ccc; \quad baa, bba.$$

The resolvability of the first three ambiguities is immediate. But in the last two cases, if we reduce the monomial in question in the two possible ways, equate the resulting expressions, and complete the reductions, we get a nontrivial equation. This turns out to be the same for both cases. Isolating a judiciously chosen term, it reads:

$$bca = abc + acb + cab + cac + cbc + 2ab + 2ac + 2cb + bc + ca. \tag{5}$$

If this is added to our list of reduction rules, then *baa* and *bba* will now reduce to unique expressions. Two new ambiguities have been created, however: *bbca* and *bcaa*. But when we analyze each of these in the same way (a full-page computation) all terms cancel. Accepting this gift-horse without question, we conclude that no more formulae need be added to our list.

The monomials irreducible under this reduction system are those words in a, b and c in which no letter occurs twice in succession, and the sequences *ba* and *bca* never occur; i.e. b never occurs (immediately or at a distance) to the left of a.

To see that this procedure eventually terminates for every element, note that in each of (1)–(5), the right-hand side is a linear combination of terms that are either of shorter length than the left-hand side, or have the same length as the left-hand side, fewer occurrences of b to the left of a, and no more total occurrences of b or a. From this it follows that the same will be true for the equations obtained by applying one of these reductions to a subword of any longer word. (The observation about total occurrences of b and a is needed for this deduction. If it failed, e.g. if the right-hand side of (5) had an *aca* term, then on reducing *bcbca*, the number of occurrences of b to the left of a would not be decreased.) In formal language, then, the partial ordering of monomials that sets $U < V$ if U is shorter than V, or of the same length but with fewer b's to the left of a's and no more a's or b's than V, is a semigroup partial ordering with descending chain condition, and compatible with our system of reductions (1)–(5). As all ambiguities of that system are resolvable, we conclude that the irreducible words form a $\mathbb{Z}$-basis for R. In particular, 2 and 3 are not zero-divisors in R, and $ab \neq 0$, so the original question is answered in the negative

Mauldon, inspired by the same problem, obtains essentially this result in [41]. Though the calculations involved are the same, his method of justifying the conclusion is not, and one can extract from it a somewhat different version of Theorem 1.2, which I will discuss in Section 11.1.

2.2. *Generalizations and Questions*

Mauldon also remarks that the indicated normal form continues to hold if the right-hand side of (4_0) is replaced by $n(a+b+c)$ for any integer n, though the reduction formulas (4) and (5), and hence the computations needed to verify the normal form, become messier. I record below the result of a still more general computation suggested by Mauldon's. (Note that in the situation described below one immediately uses (6) to eliminate the generator d, but its presence makes the initial presentation more symmetrical. Our original problem is the case $\alpha = \beta = \gamma = -1$, $\delta = 1$, $\alpha' = \beta' = \gamma' = \delta' = 0$.)

PROPOSITION 2.2.1. *Let k be a commutative ring, $\alpha,\ldots,\delta'$ elements of k, and R the k-algebra presented by generators a, b, c, d and relations*

$$a + b + c + d = 0, \tag{6}$$

$$a^2 + \alpha a + \alpha' = 0, \quad b^2 + \beta b + \beta' = 0, \quad c^2 + \gamma c + \gamma' = 0, \quad d^2 + \delta d + \delta' = 0. \tag{7}$$

Then a basis for R is given by the set of all words in a, b, c in which no letter appears twice in succession, and b never occurs to the left of a. ∎

Sketch of proof. One obtains reduction formulae for this algebra in terms of the generators a, b, c, generalizing (1)–(5). The conditions for resolvability of all ambiguities in these reductions must be a family of polynomial equations in $\alpha,\ldots,\delta'$. Hence these will hold for all values of $\alpha,\ldots,\delta'$ in all commutative rings k if they hold for all choices of these elements in a field k of characteristic 0. In that case we can make the change of variables $A = a + \alpha/2$, $B = b + \beta/2$, $C = c - (\alpha + \beta + \delta)/2$, $D = d + \delta/2$, which gives us defining relations of the same form, but with $\alpha = \beta = \delta = 0$, and this case is computationally manageable. ∎

The number of monomials of length $<n$ in the basis described in Proposition 2.2.1 is n^2. This nonexponential rate of growth implies that if R is without zero divisors, it is an Ore ring. It would be interesting to know whether R is indeed without zero divisors when the quadratic polynomials in (7) are all irreducible; whether, when it does have zero-divisors, it is a Goldie ring, and if so, what its Artinian ring of fractions looks like.

Observations and questions of the sort Mauldon makes for his rings [41, Theorem 4, and conjecture, p. 972] can also be made for rings in this wider class. As a simple example, if k is a field of characteristic 0, and $\alpha', \beta', \gamma', \delta'$ are

zero, then one can show by looking at traces that any homomorphism of R into a matrix ring over k will annihilate a, b, c if $\alpha, \beta, \gamma, \delta$ are either (i) positive rational numbers, or (ii) linearly independent over $\mathbb{Q} \subseteq k$, or most generally (iii) are simultaneously sent to positive values under some $\mathbb{Q}$-linear map $k \to \mathbb{Q}$. One may ask whether in this situation the ideal of R generated by a, b, and c is the only nonzero ideal.

3. Application: The Poincaré–Birkhoff–Witt Theorem

Let k be a commutative ring and $\mathfrak{L}$ a Lie algebra over k which is free as a k-module, on a basis X. Let us form the free associative k-algebra $k\langle X\rangle$, and identify $\mathfrak{L}$ with the k-submodule of $k\langle X\rangle$ spanned by X. Thus we have Lie brackets $[\ ,\ ]$ operating on that submodule of $k\langle X\rangle$.

Let $k[\mathfrak{L}]$ denote $k\langle X\rangle/I$, where I is the 2-sided ideal generated by all elements $ab - ba - [a, b]$ $(a, b \in \mathfrak{L})$. Given $a \in \mathfrak{L}$ we shall write a' for the image of a in $k[\mathfrak{L}]$. This ring is the *universal enveloping algebra* of $\mathfrak{L}$, i.e., it is universal among associative k-algebras R given with a k-linear map $a \mapsto a' : \mathfrak{L} \to R$ such that $[a, b]' = a'b' - b'a'$ $(a, b \in \mathfrak{L})$. (If we write $[\ ,\ ]_R$ for commutator brackets in an associative algebra R, this equation takes the form $[a', b']_R = [a, b]'$; the given generators for I likewise become $[a, b]_{k\langle X\rangle} - [a, b]$.)

Let us choose any *total ordering* $\leqslant$ on X. Note that by the k-bilinearity of the Lie brackets of $\mathfrak{L}$, the ideal I will be generated by the elements $xy - yx - [x, y]$ with $x, y \in X$, while by the antisymmetry of the bracket, such elements with $x < y$ will in fact generate the whole ideal.

Theorem 3.1 (Poincaré [47], Birkhoff [11], Witt [60].) *$k[\mathfrak{L}]$ is free as a k-module, on the basis consisting of all products $x'y' \cdots z'$ such that $x, y, \ldots, z \in X$, and $x \leqslant y \leqslant \cdots \leqslant z$.*

Proof. Let S be the reduction system on $k\langle X\rangle$ consisting of the pairs $\sigma_{xy} = (yx, xy - [x, y])$ for all $y > x$ in X. Thus, the ideal generated by the differences $W_\sigma - f_\sigma$ $(\sigma \in S)$ is precisely I, and the images in $k[\mathfrak{L}]$ of the words irreducible under S are precisely the alleged basis.

To show that this system of reductions must terminate, let us define the *misordering index* of an element $x_1 \cdots x_n \in \langle X\rangle$ as the number of pairs (i, j) such that $i < j$ but $x_i > x_j$. (For example, 0 if $x_1 \leqslant \cdots \leqslant x_n$; $n(n-1)/2$ if $x_1 > \cdots > x_n$.) Let us partially order $\langle X\rangle$ by setting $A < B$ if A is of smaller length than B, or if A is a permutation of the terms of B but has smaller misordering index. Then it is straightforward to check that $<$ is a semigroup partial ordering of $\langle X\rangle$, compatible with S, and having descending chain condition.

Note that for any $a, b \in \mathfrak{L}$,

$$ab - ba - [a, b] \in I_C \text{ for any } C \in \langle X \rangle \text{ of length } > 2. \tag{8}$$

The ambiguities of S are clearly precisely the 5-tuples $(\sigma_{zy}, \sigma_{yx}, z, y, x)$ with $z > y > x \in X$. To resolve such an ambiguity relative to $\leqslant$ we must study the element

$$r_{1\sigma_{zy}x}(zyx) - r_{z\sigma_{yx}1}(zyx) = (yzx - [y, z]x) - (zxy - z[x, y]).$$

To further reduce the term yzx, we apply first $r_{y\sigma_{zx}1}$, and then to handle the yxz which results, $r_{1\sigma_{yx}z}$. Similarly, to deal with zxy we apply $r_{1\sigma_{zx}y}$ and then $r_{x\sigma_{zy}1}$. We thus get

$$\begin{aligned}(xyz - [x, y]z - y[x, z] - [y, z]x) - (xyz - x[y, z] - [x, z]y - z[x, y]) \\ = (x[y, z] - [y, z]x) - (y[x, z] - [x, z]y) + (z[x, y] - [x, y]z).\end{aligned}$$

By (8) this is congruent modulo I_{zyx} to $[x, [y, z]] + [y, [z, x]] + [z, [x, y]]$, which is zero by the Jacobi identity in $\mathfrak{L}$. So our ambiguities are resolvable relative to $\leqslant$. Hence by Theorem 1.2, $k[\mathfrak{L}]$ has the basis indicated. ∎

The above is quite close to Birkhoff's original proof, which in fact contains the idea of Theorem 1.2, though not an explicit formulation thereof. Witt's proof looks rather different. He considers a certain action of the permutation group S_n upon the space spanned by monomials of degree $\leqslant n$. The Jacobi identity turns out to correspond to the defining relations $((i, i+1)(i+1, i+2))^3 = 1$ in a presentation of S_n in terms of the generators $(i, i+1)$. Poincaré's 1899 proof [47, Sect. III] is more or less by "brute force", and appears to have a serious gap, but it is a surprizingly early example of the idea of constructing a ring as the factor-algebra of a free associative algebra by (in effect) the ideal generated by a system of relators.

We remark that for k a field and $\mathfrak{L}, \mathfrak{L}'$ Lie algebras over k, it is an open question whether $k[\mathfrak{L}] \cong k[\mathfrak{L}']$ implies $\mathfrak{L} \cong \mathfrak{L}'$; even whether any $\mathfrak{L}$ such that $k[\mathfrak{L}]$ is a free associative algebra must be a free Lie algebra!

4. Applications: How to Take Advantage of Identities

It is not surprizing that to work out the rather arbitrary problem of Section 2 one should have to make nontrivial computations (though I did the dirty work behind the scenes and only reported the results). But one would hope that in the proof of the Poincaré–Birkhoff–Witt theorem, even the amount of "scratch-work" we went through above could be replaced by something more conceptual. Specifically, assuming one has already made the fundamental calculation showing that the commutator operation of an associative algebra satisfies the Jacobi

identity, one should be able to use this fact instead of repeating what is in essence the same calculation.

And in fact, one can. Given $z > y > x$ as in the proof of Theorem 3.1, the Jacobi identity for commutators in $k\langle X\rangle$ says that

$$[x, [y, z]_{k\langle X\rangle}]_{k\langle X\rangle} + [y, [z, x]_{k\langle X\rangle}]_{k\langle X\rangle} + [z, [x, y]_{k\langle X\rangle}]_{k\langle X\rangle} = 0. \tag{9}$$

We observe that by applying reductions from S to appropriate terms, we can reduce the left hand side of (9), first to $[x, [y, z]]_{k\langle X\rangle} + [y, [z, x]]_{k\langle X\rangle} + [z, [x, y]]_{k\langle X\rangle}$, and finally to $[x, [y, z]] + [y, [x, z]] + [z, [x, y]]$, which is zero by the Jacobi identity in $\mathfrak{L}$. (So what?) Now note that if the terms on the left-hand-side of (9) are expanded, the monomial zyx appears exactly twice, in the first term and in the third (with opposite signs of course), and that in the reductions we have performed, $r_{1\sigma_{zy}x}$ acted on the first term, and $r_{z\sigma_{yx}1}$ on the second. Further, all the other reductions we performed on the left-hand side of (9) affected monomials that were $< zyx$ under the partial ordering of Theorem 3.1, hence changed the value by members of I_{zyx}. Since we started and ended with 0, we may conclude that $r_{1\sigma_{zy}x}(zyx) - r_{z\sigma_{yx}1}(zyx) \in I_{zyx}$, as desired.

The following lemma more or less states the general principle involved here, and the first corollary covers the above application.

LEMMA 4.1. *Suppose S is a reduction system on a free algebra $k\langle X\rangle$, and "$\leqslant$" a semigroup partial ordering on $\langle X\rangle$ compatible with S. Let*

$$(\sigma, \tau, A, B, C) \tag{10}$$

be an overlap ambiguity (respectively an inclusion ambiguity) of S, and suppose

$$f_1 + \cdots + f_n = 0 \tag{11}$$

is an equation holding in $k\langle X\rangle$, such that only two terms of this equation, say f_i and f_j, involve the monomial ABC with nonzero coefficient, and in these its coefficients are units of k. Then the ambiguity (10) *is resolvable relative to $\leqslant$ if and only if*

$$f_1 + \cdots + r_{1\sigma C}(f_i) + \cdots + r_{A\tau 1}(f_j) + \cdots + f_n \in I_{ABC} \tag{12}$$

(or the analogous formula in the case of an inclusion ambiguity).

Proof. Trivial. ∎

I say this "more or less" captures our principle because it ignores the fact that when for instance we applied $r_{1\sigma_{zy}x}$ to the monomial zyx in the first term of (9), in reducing $[x, [y, z]_{k\langle X\rangle}]_{k\langle X\rangle}$ to $[x, [y, z]]_{k\langle X\rangle}$, we were simultaneously applying $r_{x\sigma_{zy}1}$ to the monomial xzy of the same term, so that we never explicitly wrote the expression corresponding to (12). So a *caveat* in applying the above lemma is to make sure that not only subsequent reductions, but also reductions performed

"together with" our first one, only affect monomials $<ABC$, under our partial ordering. When we abstract the above calculation in the next corollary, this *caveat* takes the form of the hypotheses that $zyx > yxz$, $zyz > xzy$, and in subsequent corollaries it is represented by similar order-conditions.

The first sentence of Lemma 4.1 *above will be a general hypothesis for all the corollaries in this section. We shall henceforth write* "[,]" *for the commutator operation in associative algebras*, since there will no longer be another Lie operation from which it must be distinguished.

COROLLARY 4.2. *Suppose x, y, z are elements of X, such that $zyx > yxz$ and $zyx > xzy$ under our partial ordering, and suppose S contains reductions*

$$\begin{aligned} \sigma &= (zy, yz + a), \\ \tau &= (yx, xy + b), \end{aligned} \qquad (a, b \in k\langle X\rangle).$$

Then the ambiguity (σ, τ, z, y, x) is resolvable relative to "$\leqslant$" if and only if our reduction system respects the Jacobi identity

$$[x, [y, z]] + [y, [z, x]] + [z, [x, y]] = 0$$

in the sense that

$$[x, a] + [y, [z, x]] + [z, b] \in I_{zyx}. \quad \blacksquare$$

Actually, even having the example of the Birkhoff–Witt theorem, and another to be mentioned at the end of this section, I did not see through to the principle of Lemma 4.1 until Earl Taft happily provided me with some normal form problems arising in the study of Hopf algebras in characteristic p. Let me first give the additional general results needed for these, and then the specific examples. We recall that for x an element of an associative algebra R, ad x: $R \to R$ is defined to be the map $y \mapsto [x, y] = xy - yx$.

COROLLARY 4.3. *Suppose that* char $k = p$, *a finite prime, i.e., that k is a commutative $\mathbb{Z}_p$-algebra, suppose x, y are elements of X such that $yx > xy$ (respectively $xy > yx$) and suppose S contains elements*

$$\begin{aligned} \sigma &= (yx, xy - a) \qquad (resp.\ (xy, yx + a), \\ \tau &= (x^p, b). \end{aligned}$$

Then the ambiguity $(\sigma, \tau, y, x, x^{p-1})$ (resp. $(\tau, \sigma, x^{p-1}, x, y)$) is resolvable relative to $\leqslant$ if and only if our reduction system respects the identity

$$(\text{ad } x)^p(y) = (\text{ad } x^p)(y) \qquad [31, \text{Chap. V, (60), p. 186}]$$

in the sense that

$$(\text{ad } x)^{p-1}(a) - (\text{ad } b)(y) \in I_{yx^p} \text{ (resp. } I_{x^py}). \quad \blacksquare$$

COROLLARY 4.4. *Suppose x is an element of X and n a positive integer such that S contains an element*

$$\sigma = (x^n, a).$$

Then all the ambiguities $(\sigma, \sigma, x^i, x^{n-i}, x^i)$ are resolvable relative to $\leqslant$ if and only if our reduction system respects the identity

$$[x^n, x] = 0$$

in the sense that

$$[a, x] \in I_{x^{n+1}}. \quad \blacksquare$$

The reader can now establish the two results Taft asked for. All algebras are over a field k of characteristic p:

EXERCISE 4.5. Suppose $p \neq 2$. Then the algebra presented by generators x, y, z and relations

$$[x, y] = x, \quad [y, z] = -z, \quad [x, z] = x^2/2, \quad x^p = 0, \quad y^p = y, \quad z^p = 0, \qquad \text{(Taft and Wilson [55])}$$

has dimension p^3, with basis $\{x^a y^b z^c \mid 0 \leqslant a, b, c < p\}$. (First find a formula for the nth power of the derivation on $k[x]$ taking x to $x^2/2$.)

The following simpler-looking case turned out to be harder:

EXERCISE 4.6. The algebra presented by generators x, y and relations

$$[x, y] = y^2 - y, \quad x^p = x, \quad y^p = 1 \qquad \text{(Radford [56, p. 158])}$$

has dimension p^2, with basis $\{x^a y^b \mid 0 \leqslant a, b < p\}$. (If D: $k[y] \to k[y]$ is the derivation such that $D(y) = y^2 - y$, there is not a simple formula for $D^n(y)$. But if we define two other derivations by $D_1(y) = y^2$, $D_2(y) = -y$, then $D = D_1 + D_2$. Observing that $[D_1, D_2] = D_1$, and using the formula for the pth power of a sum of derivations in characteristic p [31, Sect. 5.7], you can evaluate $D^p(y)$, and thus make the desired calculations.)

EXERCISE 4.7. Prove the analog for p-Lie algebras of the Poincaré–Birkhoff–Witt theorem [31, Theorem 5.12].

Finally, we have the following application of Lemma 4.1, which is relevant to Cohn's and my work concerning algebras R with systems of matrices over R having universal properties. For example, if m and n are positive integers and k a commutative ring, then the k-algebra R with a universal module-isomorphism ξ: $R^m \cong R^n$ may be presented by $2mn$ generators, mn of which are the entries x_{ij} of the $m \times n$ matrix x representing ξ, and the other mn, the entries y_{ij} of the

inverse matrix y, together with defining relations constituting the matrix equations $xy = I_m$, $yx = I_n$, namely,

$$\sum_i x_{hi}y_{ij} = \delta_{hj}, \qquad \sum_i y_{hi}x_{ij} = \delta_{hj}.$$

In the following corollary, the families of elements $((x_{gh}))$,..., $((v_{hj}))$ should likewise be thought of as the entries of rectangular matrices x, y, z, u, v. (For a general discussion of such universal-matrix constructions see [5, Introduction]. For normal-form results see [5, Sect. 9, in particular pp. 62–63]; cf. [17, 19, 37].)

COROLLARY 4.8. *Suppose p, m, n, q are positive integers, and that X contains elements x_{gh}, y_{hi}, z_{ij} $(g \leqslant p, h \leqslant m, i \leqslant n, j \leqslant q)$. Suppose $\mu \leqslant m$ is a positive integer such that $x_{g\mu}y_{\mu i} > x_{gh}y_{hi}$ for all $h \neq \mu$, and $\nu \leqslant n$ a positive integer such that $y_{h\nu}z_{\nu j} > y_{hi}z_{ij}$ for all $i \neq \nu$, and that S contains reductions*

$$\begin{aligned} \sigma_{gi} &= \left(x_{g\mu}y_{\mu i}, -\sum_{h\neq\mu} x_{gh}y_{hi} + u_{gi}\right) \qquad (g \leqslant p, i \leqslant n), \\ \tau_{hj} &= \left(y_{h\nu}z_{\nu j}, -\sum_{i\neq\nu} y_{hi}z_{ij} + v_{hj}\right) \qquad (h \leqslant m, j \leqslant q). \end{aligned} \tag{13}$$

where $u_{gi}, v_{hj} \in k\langle X\rangle$. (This system of reductions represents matrix relations $xy = u$, $yz = v$, with the μth, respectively the νth, summand of each entry of the product-matrix taken as the word to be reduced.)

Then the pq ambiguities $(\sigma_{gv}, \tau_{uj}, x_{gu}, y_{uv}, z_{vj})$ $(g \leqslant p; j \leqslant q)$ are resolvable relative to $\leqslant$ if and only if the reduction system S respects the associative identity for matrix multiplication

$$(xy)z = x(yz)$$

in the sense that

$$uz - xv \in ((I_{x_{g\mu}y_{\mu\nu}z_{\nu j}})). \quad \blacksquare \tag{14}$$

For normal-form calculations based similarly on the associative identity for multiplication of formal power series or formal Laurent series, see [75].

Remark 4.9. If in Lemma 4.1 we add the hypothesis that all monomials $<ABC$ are reduction-unique, which will generally be known to be true in any useful application of that Lemma, then the hypothesis that the two nonzero coefficients of ABC in (11) be units can be weakened to require only that they be non-zero-divisors in k. For when all monomials $<ABC$ are reduction-unique, I_{ABC} can be characterized as $\mathrm{Ker}(r_S \mid \text{span of } \{\text{monomials} < ABC\})$; so if we write α and $-\alpha$ for the two coefficients of ABC in (11), then (12) implies that $\alpha(f_\sigma C - Af_\tau) \in I_{ABC}$, whence $\alpha r_S(f_\sigma C - Af_\tau) = 0$, whence $r_S(f_\sigma C - Af_\tau) = 0$, so the given ambiguity is in fact resolvable.

Throughout this section we have based our computations on *identities* for associative algebras. There may also be situations in which one would like to use in the same way some *equation* known to hold in the ring R one is working with. One can indeed do this, *provided that* the equation in question arises from "lower" applications of the reduction rules, as indicated in (15) below. Though the formal statement is obviously equivalent to Lemma 4.1 (which is itself obvious), I record it as a reminder of the possibility of using this trick.

COROLLARY 4.10. *Lemma* 4.1 *remains true if* (11) *is weakened to*

$$f_1 + \cdots + f_n \in I_{ABC}\,. \quad \blacksquare \tag{15}$$

5. OBSERVATIONS ON SOME PRACTICAL QUESTIONS

5.1. The reader may wonder why we have seen no cases of inclusion ambiguities among our examples. This is because inclusion ambiguities are, in a sense, always avoidable.

Suppose S is a reduction system for a free algebra $k\langle X\rangle$. Let us construct a subset $S' \subseteq S$ by (i) dropping all $\sigma \in S$ such that W_σ contains a *proper* subword of the form W_τ ($\tau \in S$), and (ii) whenever more than one element, $\sigma_1, \sigma_2, \ldots \in S$ act on the same monomial (i.e., $W_{\sigma_1} = W_{\sigma_2} = \cdots$) dropping all but one of the σ_i from S. Then S' will have no inclusion ambiguities, and will have the property that a member of $k\langle X\rangle$ is reducible under S' if and only if it is reducible under S. But from this it follows that if $a \in k\langle X\rangle$ is reduction-unique under S, then it is reduction-unique under S' and $r_{S'}(a) = r_S(a)$. Hence if S is such that *every* element of $k\langle X\rangle$ is reduction-unique under it, then S' has the same property, and $r_{S'} = r_S$. Thus S', which has no inclusion ambituities, defines the same ring and the same canonical form as S.

5.2. The example of Section 2 (idempotents $a, b, c, \ldots$) was the only one of our applications where we saw how one *finds* a normal form for a given ring. When our original reduction formulae (1)–(4) turned out to have nonresolvable ambiguities, these led to new equations (k-linear relations holding in R among monomials irreducible under our existing reductions) which we made into new reduction formulae.

Note that in this process, we had to make a choice of which term of each equation to isolate as our "W_σ." Some observations on the consequences of this choice for Eq. (5) in the above-mentioned example are instructive: If we had used *acb* rather than *bca*, this would have resulted in four rather than two new ambiguities. Now an ambiguity means that more than one reduction can be applied to some monomial, hence whether resolvable or not, it represents a kind of inefficiency in our reduction system. By a counting argument one can deduce that there must necessarily be a greater number of monomials of length

$\leqslant 4$ *irreducible* under this alternative system than under the more efficient one. But the rank of the k-submodule of R spanned by all monomials of degree $\leqslant 4$ in a, b and c is invariant. It follows that this alternative reduction system would not have given unique normal forms for the elements of R. Hence it must have irresolvable ambiguities, and more reductions would have to be added before a system yielding a normal form was obtained. This suggests that a good heuristic principle is to choose reduction formulae at each step so as to minimize the number of ambiguities resulting.

Still other choices of the term to isolate in (5), especially various choices of terms of length 2 instead of 3, would have led to non-terminating reduction procedures; so this is another consideration that must be borne in mind.

EXERCISE 5.2.1. Examine for termination each of the following singleton reduction-systems on $k\langle x, y\rangle$: $\{(x^2y, yx)\}$, $\{(yx, x^2y)\}$, $\{(x^2y^2, yx)\}$, $\{(yx, x^2y^2)\}$.

5.3. Will an algebra R generated by a set X in general have a normal form of the sort we have been studying? If k is a field, the answer is yes:

Recall that a total ordering with descending chain condition is called a well-ordering, and that it is easy to find semigroup well-orderings on a free semigroup $\langle X\rangle$. For example, well-order X (especially easy if it is finite) and for $A, B \in \langle X\rangle$, put $A < B$ if A is shorter than B, or has the same length but precedes it lexicographically.

So suppose that R is an algebra over a field k, generated by a set X, and let "$\leqslant$" be a semigroup well-ordering on the free semigroup (with 1) $\langle X\rangle$. Let $\langle X\rangle_{\text{irr}}$ denote the set of elements $A \in \langle X\rangle$ whose images in R are not equal to k-linear combinations of the images of elements $B < A$. From the fact that "$\leqslant$" is a semigroup ordering, it follows that if $A \in \langle X\rangle_{\text{irr}}$, then all subwords of A also belong to $\langle X\rangle_{\text{irr}}$. Now let $Z \subseteq \langle X\rangle$ denote the set of all monomials which do *not* belong to $\langle X\rangle_{\text{irr}}$, but whose proper subwords all do belong to this set. For each $W \in Z$, let us write the image of W in R as a k-linear combination of images of elements $< W$; this is the image of a certain element $f \in k\langle X\rangle$. Let S be the set of pairs (W, f) so obtained. Then it is easy to verify that S is a reduction system (without inclusion ambiguities), which is compatible with "$\leqslant$," and reduces elements $a \in k\langle X\rangle$ to unique elements of $k\langle X\rangle_{\text{irr}}$ representing the image of a in R, and hence giving a normal form for elements of R.

Likewise, if we are given generators X and relations for a k-algebra R, we may construct a reduction system for R by choosing a semigroup well-ordering "$\leqslant$" for $\langle X\rangle$, writing each relation as a formula which reduces its $\leqslant$-maximal term to a linear combination of the others (k is still assumed a field!) and systematically resolving any ambiguities by making the new equations which they yield into new reduction formulae in the same way.

The existence of this algorithm does not contradict the various results on the unsolvability of word problems. Those results say that the classes of relations

satisfied in certain finitely presented objects are not recursive; but trivially, they are recursively enumerable, and the process described above is just a systematic and perhaps reasonably efficient way of enumerating them in the case of associative algebras. Note that if R is a finitely presented k-algebra with unsolvable word problem, any reduction system S for R satisfying the conditions of Theorem 1.2 must be infinite in fact, nonrecursive.

Likewise, if R is not finitely related, any reduction system giving a canonical form for elements of R must be infinite. But such infinite systems need not be unpleasant. The reader might find it interesting to work out a reduction system for the subalgebra of a free algebra $k\langle x, y\rangle$ generated by the four elements x, xy, yx, yy (cf. [10] for some interesting results and an open question related to this algebra).

Note that if we are given a finite presentation of a k-algebra R, the technique described above for finding a reduction system is not guaranteed to yield a finite (or in some other sense convenient) reduction system even if one exists. That, presumably, must remain an art. The remarks at the end of the previous subsection may serve as first guidelines.

Note in this connection that, though it is known that any partial ordering with descending chain condition on a set X can be strengthened to a well-ordering, the corresponding statement for semigroup partial orderings on $\langle X\rangle$ is false. For example, let $X = \{u, v, x, y\}$. Then the semigroup preorder on $\langle X\rangle$ "generated" by the relations $xu \geqslant yu$, $yv \geqslant xv$ is easily shown to be a partial order with descending chain condition. But it clearly cannot be extended to a semigroup partial ordering satisfying either $x \geqslant y$ or $y \geqslant x$. So a ring R might conceivably have a finite or recursive reduction-system compatible with some semigroup partial ordering on $\langle X\rangle$, but none compatible with a total ordering, hence none obtainable as described in this section.

5.4. Finally, let us ask: If S is a reduction system on a free algebra $k\langle X\rangle$, under which all elements are known to be reduction-finite, will there in general exist a semigroup partial order $\leqslant$ with descending chain condition compatible with S?

Given S, let $<$ denote the least transitive relation on $\langle X\rangle$ such that $C < D$ whenever C occurs with nonzero coefficient in $r_{A\sigma B}(D)$, for some reduction $r_{A\sigma B}$ acting nontrivially on D. It is clear that "$<$" respects the semigroup structure of $\langle X\rangle$. Now if k has no zero divisors, one can show with a little ingenuity that all elements of $k\langle X\rangle$ are reduction-finite under S if and only if $<$ is antireflexive and has descending chain condition, which means precisely that the relation "$A < B$ or $A = B$" is a partial ordering with descending chain condition. By construction, it is compatible with S. (In fact it is the minimal partial ordering compatible with S.) On the other hand, if k has zero-divisors $ab = 0$, consider the free algebra $k\langle u, x, y\rangle$ and the reduction system consisting of the pairs (ux, auy) and (yu, bxu). Under any partial ordering "$\leqslant$" compatible

with S one would have $uxu > uyu > uxu$, a contradiction. Nevertheless, all elements of $k\langle X\rangle$ are reduction-finite under S; in particular, one has the reductions $uxu \mapsto a\,uyu \mapsto ab\,uxu = 0$. (In fact, every element of $k\langle X\rangle$ is reduction-*unique*, and as in Theorem 1.2, this yields a k-basis of monomials for the ring defined.)

6. Generalization: Bimodule-Structures of k-Rings

Not all ring-theoretic constructions for which one has normal-form results are algebras over the given base ring k. As a first step away from that case, note that constructions of such algebras often have "twisted" analogs, in which indeterminates x are introduced which rather than commuting with k, satisfy relations such as

$$cx = xc^{\alpha_x} \qquad (c \in k), \tag{16}$$

where α_x is a ring-endomorphism of k. In this situation it is also natural to drop the assumption that k is commutative. We recall that for k an arbitrary ring with 1, a ring R given with a (unital) homomorphism $k \to R$ is called a *k-ring* (generalizing the concept of *k-algebra*.)

Can we generalize Theorem 1.2 to the sort of construction indicated above? It is evident that some compatibility conditions are needed. For example, if we have a reduction $(xy, z) \in S$ then we will expect the associated endomorphisms to satisfy

$$\alpha_x\alpha_y = \alpha_z\,. \tag{17}$$

If we think of (16) as describing a *k-bimodule* structure on the *right k-module* xk, then (17) says that the reduction $xy \mapsto z$ should induce a bimodule homomorphism

$$(xk) \otimes_k (yk) \to zk.$$

This suggests the following more general development.

Let k be an associative ring with unit, and $(M_x)_X$ an arbitrary family of k-bimodules indexed by a set X. For every $A = x_1 \cdots x_n \in \langle X\rangle$, let P_A denote the k-bimodule $M_{x_1} \otimes \cdots \otimes M_{x_n}$. Here we understand $P_1 = k$, and $P_x = M_x$ $(x \in X)$. Then $\oplus_{\langle X\rangle} P_A$ will be the tensor ring on the k-bimodule $M = \oplus_X M_x$, and we will denote it $k\langle M\rangle$. By a reduction system for $k\langle M\rangle$ we shall understand a set S of pairs $\sigma = (W_\sigma, f_\sigma)$, where $W_\sigma \in \langle X\rangle$, and the second component is a *k-bimodule homomorphism* $f_\sigma\colon P_{W_\sigma} \to k\langle M\rangle$. For any $A, B \in \langle X\rangle$, $\sigma \in S$, we define $r_{A\sigma B}\colon k\langle M\rangle \to k\langle M\rangle$ to be the k-bimodule homomorphism which acts on $P_{AW_\sigma B} = P_A \otimes P_{W_\sigma} \otimes P_B$ by the map $P_A \otimes f_\sigma \otimes P_B$, and is the identity on all other P_C $(C \in \langle X\rangle - \{AW_\sigma B\})$. The set of elements of $\langle X\rangle$ containing no subwords of the form W_σ $(\sigma \in S)$ will be denoted $\langle X\rangle_{\mathrm{irr}}$, and we define the submodule $k\langle M\rangle_{\mathrm{irr}}$ of irreducible elements of $k\langle M\rangle$ as $\oplus_{A \in \langle X\rangle_{\mathrm{irr}}} P_A$.

One can now parallel the development in Section 1 quite closely, so I shall only point out the differences, though the reader may wish to think things through in detail.

Reduction-finiteness and reduction-uniqueness of elements of $k\langle M\rangle$, and the map r_S, are defined just as in Section 1. Lemma 1.1 carries over, with the reference to "monomials A, B, C occurring in a, b, c with nonzero coefficients" replaced by "the homogeneous components a_A, b_B, c_C of a, b, c in any of the summands P_A, P_B, P_C of $k\langle M\rangle$." An ambiguity $(\sigma, \tau, A, B, C) \in S^2 \times \langle X\rangle^3$ is defined precisely as in Section 1. An overlap ambiguity (σ, τ, A, B, C) will be called resolvable if *for all* $d \in P_{ABC}$, the elements $r_{1\sigma C}(d)$ and $r_{A\tau 1}(d)$ can be reduced to a common element; similarly for inclusion ambiguities.

We call a semigroup partial ordering $\leqslant$ on $\langle X\rangle$ compatible with S if $f_\sigma(P_{W_\sigma}) \subseteq \oplus_{A<W_\sigma} P_A$ for all $\sigma \in S$. We make the obvious definitions of the 2-sided ideal $I \subseteq k\langle M\rangle$ and of the sub-k-bimodules I_A ($A \in \langle X\rangle$), and we say that an overlap (respectively inclusion) ambiguity (σ, τ, A, B, C) is resolvable *relative* to $\leqslant$ if $r_{1\sigma C} - r_{A\tau 1}$ carries P_{ABC} into I_{ABC}.

We now get, by essentially the same argument, the following analog (and generalization) of Theorem 1.2. I have pared the statement of the conclusion down to essentials to avoid being repetitious.

THEOREM 6.1. *Let k be an associative ring with 1, $k\langle M\rangle$ the tensor ring over k on the direct sum M of a family $(M_x)_{x\in X}$ of k-bimodules, S a reduction system for $k\langle M\rangle$, I the ideal of $k\langle M\rangle$ generated by $\{a - r_\sigma(a) \mid \sigma \in S, a \in P_{W_\sigma}\}$, and $\leqslant$ a semigroup partial ordering of $\langle X\rangle$ compatible with S and having descending chain condition.*

Then the map $k\langle M\rangle_{\mathrm{irr}} \to R = k\langle M\rangle/I$ is a k-bimodule isomorphism if and only if all ambiguities of S are resolvable (relative to $\leqslant$). ∎

This result is not a canonical form statement in the same concrete sense as Theorem 1.2 of course, except in situations where we can get nice canonical forms for the P_A's. But such structure results can be useful in the same way as canonical form statements. We shall see an application in Section 8.

Recall that if k-bimodules M_1, M_2 are free as *right* k-modules, on bases Z_1, Z_2, then $M_1 \otimes_k M_2$ is right free on the basis $\{z_1 \otimes z_2 \mid z_1 \in Z_1, z_2 \in Z_2\}$. Hence under appropriate hypotheses the above Theorem yields right k-bases for k-rings R.

COROLLARY 6.2. *Suppose the equivalent conditions of the preceding Theorem are satisfied by $k\langle M\rangle$, S, $\leqslant$, and that for each $x \in X$, M_x is free as a right k-module on a basis $Z(x)$, where we assume the $Z(x)$'s pairwise disjoint. Then R is free as a right k-module on the basis $\{z_1 \cdots z_n \mid n \geqslant 0, z_i \in Z(x_i), x_1 \cdots x_n \in \langle X\rangle_{\mathrm{irr}}\}$.* ∎

Remark. It is curious that an arbitrary system of direct sum decompositions of the M_x as right k-modules does not, similarly, induce a right-k-direct-sum

decomposition of each P_A. The key to the situation seems to be the following: If we fix a right k-module M (possibly a bimodule, but we will not be concerned with its left module structure) and look at the functor $M \otimes -$: k-bimodules $\to$ right k-modules, then we find that for a k-bimodule N, the right k-module $M \otimes N$ is not, in general functorial in (or even determined up to isomorphism by) the *right* k-module structure of N. That is, a right k-module homomorphism $N \to N'$ between such bimodules does not induce a homomorphism $M \otimes N \to M \otimes N'$; so in particular a right direct sum decomposition of N will not induce such a decomposition of $M \otimes N$. An exception to this is when the right module M is free. Then $M \otimes -$ can be extended from a functor on bimodules to a functor on right modules. But this is still not canonical; it depends on the choice of basis. More generally, one can do this if M is represented as $A \otimes_{\mathbb{Z}} k$, where A is an Abelian group.

7. Generalization: Right Bases for k-Rings

There are still some very simple ring constructions with normal forms that are not covered by Theorem 6.1. For example, there is a "twisted polynomial ring" construction in which one adjoins to the base-ring k an element x satisfying not (16) but

$$ax = xa + a^{\partial},$$

where ∂: $k \to k$ is a derivation; or more generally

$$ax = xa^{\alpha} + a^{\partial}, \tag{18}$$

where α is an endomorphism of k, and ∂ an α-derivation (an additive map satisfying $(ab)^{\partial} = ab^{\partial} + a^{\partial}b^{\alpha}$). Here, though the $x^n k$ are not in general sub-bimodules of R, R is still free as a right k-module on $\{1, x, x^2, \ldots\}$ [44]. To handle this case, it seems we need a version of the Diamond Lemma that gives right k-bases, though perhaps not saying much about k-bimodule structure. We shall now trick Theorem 6.1 into giving us such a result. The trick can probably be used to get a more general statement than I obtain below, but I leave such extension to the reader who encounters the need for it, or is inspired with an elegant generalization.

Let k be an associative ring with 1, and X a set, and let $\langle X \rangle k$ denote the free *right* k-module on the set $\langle X \rangle$. Suppose we are given the following data:

> For each $x \in X$, an abelian group homomorphism φ_x: $k \to \langle X \rangle k$, which carries 1 to x. (For $a \in k$, $\varphi_x(a)$ represents the intended expression for ax as a right linear combination of monomials in the ring we are constructing.) (19)
>
> A family $S \subseteq \langle X \rangle \times \langle X \rangle k$ of pairs $\sigma = (W_\sigma, f_\sigma)$, thought of as reduction formulae for monomials, as in Section 1. (20)

We again define $\langle X\rangle_{\mathrm{irr}}$ to be the subset of all elements of $\langle X\rangle$ which contain no subword W_σ ($\sigma \in S$); the submodule of $\langle X\rangle k$ that these span will be called $\langle X\rangle_{\mathrm{irr}}k$.

We shall now construct from the above data a reduction system of the sort used in Theorem 7.1, but with base-ring the ring $\mathbb{Z}$ of integers. The index-set will be $X \cup \{\kappa\}$, i.e. X with a new symbol κ adjoined. For each $x \in X$ we take for M_x a free abelian group $x\mathbb{Z}$ on the one generator x, while we take M_κ to be the additive group of the ring k. Note that for every $A = x_1 \cdots x_n \in \langle X\rangle$, the abelian group $P_A = (x_1\mathbb{Z}) \otimes_{\mathbb{Z}} \cdots \otimes_{\mathbb{Z}} (x_n\mathbb{Z})$ is again free of rank 1; we will denote it $A\mathbb{Z}$. For A as above we can also form $A = x_1 \cdots x_n\kappa$, and we see that $P_{A\kappa}$ will be a free *right k-module* of rank 1; we shall denote this Ak. Thus we identify the free right k-module $\langle X\rangle k$ with a certain subgroup of $\mathbb{Z}\langle M\rangle$. (There are other elements in $\langle X \cup \{\kappa\}\rangle$, but we set up no special notation for these.)

The reduction system S' for $\mathbb{Z}\langle M\rangle$ is now defined to consist of:

> The reduction $\sigma_k = (\kappa\kappa, \mu)$, where $\mu\colon P_{\kappa\kappa} = k \otimes_{\mathbb{Z}} k \to k = M_\kappa \subseteq \mathbb{Z}\langle M\rangle$ is the map induced by the multiplication of k. (21)

> For each $x \in X$, the reduction $\sigma_x = (\kappa x, \varphi'_x)$, where $\varphi'_x\colon k \otimes_{\mathbb{Z}} (x\mathbb{Z}) \to \mathbb{Z}\langle M\rangle$ is defined by $\varphi'_x(a \otimes x) = \varphi_x(a) \in \langle X\rangle k \subseteq \mathbb{Z}\langle M\rangle$ (cf. (19)). (22)

> For each $\sigma = (W_\sigma, f_\sigma)$ in our original reduction system S, the reduction $\sigma' = (W_\sigma, f'_\sigma)$, where $f'_\sigma\colon W_\sigma\mathbb{Z} \to \mathbb{Z}\langle M\rangle$ is defined to take the generator W_σ to $f_\sigma \in \langle X\rangle k \subseteq \mathbb{Z}\langle M\rangle$. (23)

We now want a semigroup partial ordering "$\leqslant$" with descending chain condition on $\langle X \cup \{\kappa\}\rangle$ compatible with S'. We might simply assume the existence of such a "$\leqslant$"; but it is possible to describe a fairly general construction for such a partial order in terms of structure closer to our original data. Suppose we have a *semigroup preorder* $\leqslant_0$ on $\langle X\rangle$, that is, a reflexive transitive relation closed under right and left multiplications. We shall write $A <_0 B$ if $A \leqslant_0 B$ holds, but not $B \leqslant_0 A$, and assume $\leqslant_0$ has descending chain condition, that is, that there are no infinite chains $A >_0 B >_0 \cdots$ in $\langle X\rangle$. Let us write $A =_0 B$ if $A \leqslant_0 B$ and $B \leqslant_0 A$. The preorder $\leqslant_0$ will be called *consistent with S* if for all $\sigma \in S$, one has $f_\sigma \in \bigoplus_{A <_0 W_\sigma} Ak$, and *consistent with* $(\varphi_x)_{x\in X}$ if for all $x \in X$, $a \in k$, one has $\varphi_x(a) \in \bigoplus_{A \leqslant_0 x} Ak$. (Note the weaker nature of this condition.) Assuming these consistency conditions, we now define a partial order "$\leqslant$" on $\langle X \cup \{\kappa\}\rangle$ as follows. For $A \in \langle X \cup \{\kappa\}\rangle$, let A' denote the word obtained by deleting all κ's from A, let $n(A)$ denote the number of κ's in A, and let $h(A)$ denote the $n(A)$-tuple of nonnegative integers: (number of letters after the last κ in A,..., number of letters after the first κ in A). Then we shall write $A \leqslant B$ if either

(i) $A' <_0 B'$, or

(ii) $A' =_0 B'$ and $n(A) < n(B)$, or

(iii) $A' =_0 B'$, $n(A) = n(B)$, but $h(A) < h(B)$ under lexicographic ordering (i.e., $h(A) \neq h(B)$, and the first nonzero term of $h(B) - h(A)$ is positive), or

(iv) $A = B$.

That "$\leqslant$" has the required properties is now straightforward to verify. Hence if all ambiguities of S' are resolvable relative to $\leqslant$, Theorem 6.1 gives us a normal form for the ring R' having the elements of X and the elements of k as generators, and the relations given by S'.

But this R' is not quite the ring we want! The set $\langle X \cup \{\kappa\}\rangle_{\text{irr}}$ is precisely $\langle X\rangle_{\text{irr}} \cup \langle X\rangle_{\text{irr}}\kappa$, thus $\mathbb{Z}\langle M\rangle_{\text{irr}} \cong R'$ is the direct sum of $\langle X\rangle_{\text{irr}}k$ *and* a $\mathbb{Z}$-free part spanned by "bare" monomials. But in fact, it is easy to verify that the summand isomorphic to $\langle X\rangle_{\text{irr}}k$ will form a subring $R \subseteq R'$ (with a different unit: $1_k \in k = M_\kappa$ rather than $1_{\langle X\rangle} \in P_1$), and that *this* is the k-ring defined by the intended generators and relations. Thus we get:

PROPOSITION 7.1. *Let k be a ring, X a set, $(\varphi_x\colon k \to \langle X\rangle k)_{x\in X}$ a family of additive-group homomorphisms satisfying $\varphi_x(1) = x$, and S a subset of $\langle X\rangle \times \langle X\rangle k$. Let M be the additive group $k \oplus (\oplus_X x\mathbb{Z})$, and S' the reduction system on $\mathbb{Z}\langle M\rangle$ constructed above* ((21)–(23)). *Suppose $\leqslant_0$ is a semigroup preordering on $\langle X\rangle$ with descending chain condition consistent with S and (φ_x), and $\leqslant$ the induced partial ordering on $\langle X \cup \{\kappa\}\rangle$ (or more generally, let $\leqslant$ be any semigroup partial ordering on $\langle X \cup \{\kappa\}\rangle$ consistent with S'). Then the following are equivalent:*

(a) *All ambiguities of S' are resolvable (with respect to $\leqslant$).*

(b) *The k-ring R defined by the generating set X and the relations*

$$ax = \varphi_x(a) \qquad (a \in k, x \in X),$$
$$W_\sigma = f_\sigma \qquad (\sigma \in S)$$

has for right k-basis the set $\langle X\rangle_{\text{irr}}$ of monomials in X irreducible under S. ∎

Note that in a reduction system S' constructed as above, there will be four kinds of ambiguities. There are those of the form $(\sigma', \tau', A, B, C)$ $(\sigma, \tau \in S)$ arising from overlap or inclusion ambiguities of S. Then, for each $\sigma \in S$, if we write $W_\sigma = xA$ $(x \in X, A \in \langle X\rangle)$ we have an ambiguity $(\sigma_x, \sigma', \kappa, x, A)$. These test the compatibility between S and the φ_x's (cf. 17)). Third, for each $x \in X$ we have an ambiguity $(\sigma_k, \sigma_x, \kappa, \kappa, x)$. These test the consistency of the left module structure on $\langle X\rangle k$ determined by the φ_x's. Finally, there is the ambiguity $(\sigma_k, \sigma_k, \kappa, \kappa, \kappa)$, but this just tests the associativity of the multiplication of k, and so is automatically resolvable since k was assumed a ring.

If all this seems very formidable, the reader might try the easy case of k an arbitrary ring, $X = \{x\}$ a singleton, S empty, and $\varphi_x(a)$ of the form $xa^\alpha + a^\partial$ (α and ∂ linear maps), and show that necessary and sufficient conditions for the

unique nontrivial ambiguity of S', $(\sigma_k, \sigma_x, \kappa, \kappa, x)$ to be resolvable are that α be an endomorphism of k, and ∂ an α-derivation, as in (18) above, thus recovering a result of Ore [44, Chap. 1]. There will be another application in the next section.

8. Applications: Coproducts of k-Rings

Let k be an associative ring with 1, and $(R_\lambda)_{\lambda\in\Lambda}$ a family of unital k-rings. By the *coproduct* $\coprod_k R_\lambda$ of these rings over k is meant the k-ring R universal for having a k-ring homomorphism of each R_λ into it. Such a k-ring will always exist, by universal algebra [16, Theorem III.6.1].

(The coproduct is commonly called the "*free product* of the R_λ with amalgamation of k" *if* each R_λ is mapped one-to-one into R, and if their images in R are disjoint except for the common image of k. As I noted in [4, p. 2 ftn.], I consider it preferable to use the term coproduct, and describe the above situation by saying that in the given case the coproduct R is faithful in each R_λ, and "separating".)

Corollary 8.1 to Theorem 6.1. (Cf. Stallings [54].) *Let $(R_\lambda)_{\lambda\in\Lambda}$ be a family of faithful k-rings, in each of which k forms a direct summand as a k-bimodule: $R_\lambda = k \oplus M_\lambda$. Then the coproduct $R = \coprod_k R_\lambda$ is isomorphic as a k-bimodule to the direct sum of all tensor products $M_{\lambda_1} \otimes_k \cdots \otimes_k M_{\lambda_n}$, where $n \geqslant 0$, and $\lambda_1, \ldots, \lambda_n \in \Lambda$, with no two successive λ_i's equal; these summands representing the products in R of the images of the indicated k-subbimodules of the R_λ.*

Proof. For each $\lambda \in \Lambda$, let m_λ: $M_\lambda \otimes_k M_\lambda \to k \oplus M_\lambda = R_\lambda$ be the k-bimodule homomorphism induced by the multiplication of R_λ. Then each R_λ may be presented as in Theorem 6.1, using the singleton index-set $\{\lambda\}$, the one generating k-bimodule M_λ, and the reduction system consisting of the one element $\sigma_\lambda = (\lambda\lambda, m_\lambda)$. A compatible ordering with descending chain condition on the semigroup $\langle\lambda\rangle$ is given by $1 < \lambda < \lambda^2 < \cdots$. Hence the one ambiguity of this system, $(\sigma_\lambda, \sigma_\lambda, \lambda, \lambda, \lambda)$ is resolvable by Theorem 6.1(b) $\Rightarrow$ (a).

Now consider the k-ring presented by the index-set Λ, the family of k-bimodules $(M_\lambda)_{\lambda\in\Lambda}$, and the reduction system $S = \{\sigma_\lambda \mid \lambda \in \Lambda\}$. As we have merely brought together generators and relations for the separate k-rings R_λ, this k-ring must be the coproduct $R = \coprod_k R_\lambda$. We note that S will have no ambiguities but the $(\sigma_\lambda, \sigma_\lambda, \lambda, \lambda, \lambda)$ which appeared before, and which we know are all resolvable. If we partially order the free semigroup $\langle\Lambda\rangle$ by the relation "is longer than," all the hypotheses of Theorem 6.1 are satisfied, and we conclude that R is the direct sum of those of the tensor products of M_λ associated with irreducible members of $\langle\Lambda\rangle$, that is words with no two consecutive letters equal. ∎

COROLLARY 8.2 to Theorem 7.1. (Cohn, a case of [14, Theorems 4.4, 4.6]. See also [4].) *Let $(R_\lambda)_{\lambda\in\Lambda}$ be a family of k-rings, each of which is free as a right k-module on a basis of the form $B_\lambda \cup \{1\}$ (where the B_λ's will be assumed pairwise disjoint for notational purposes). Then the coproduct $R = \coprod_k R_\lambda$ is free as a right k-module on the basis consisting of all products $b_1 \cdots b_n$ such that each b_i lies in some B_{λ_i}, and no two consecutive λ_i are equal.*

Proof. For each $\lambda \in \Lambda$ we can write down systems S_λ, $(\varphi_b)_{b\in B_\lambda}$ of the sort considered in Proposition 7.1 presenting R_λ as a k-ring generated by B_λ. Here S_λ will contain for each $b, b' \in B_\lambda$ a reduction $(bb', f_{bb'})$, where $f_{bb'}$ is the expression for the product $bb' \in R_\lambda$ as a right k-linear combination of the elements of $B_\lambda \cup \{1\} \subseteq \langle B_\lambda\rangle$; the φ_b are likewise defined in the obvious way. The argument now exactly parallels that of the preceding corollary. For our preorder "$\leqslant_0$" we use the relation "length $A \leqslant$ length B." ∎

However, these results can also be proved easily without the Diamond Lemma (cf. [4]). The reason is discussed in Section 11.2 below.

Remarks. In the above two cases, we see that the coproduct ring R will be separating, will be faithful in all R_λ, and will satisfy the same conditions as a k-ring that we assumed for the R_λ.

In contrast, if we form the coproduct of one k-ring satisfying the hypotheses of the first corollary and one satisfying the hypotheses of the second, the result can be a disaster. For instance, let $k = \mathbb{Z}[t]$, and consider the k-rings $R_1 = k\langle x \mid tx = 1\rangle$, $R_2 = k\langle y \mid yt = 0\rangle$. The first is right-free over k on the basis $\langle x\rangle = \{1, x, x^2, \ldots\}$, the second has k as a k-bimodule direct summand $R_2 = k \oplus R_2 y$ (and what is more, it is *left*-free over k, on the basis $\langle y\rangle$). But in the coproduct, $y = ytx = 0x = 0$, so this is not faithful in R_2. If we also adjoin to R_2 a generator z and the relation $yz = 1$, the same properties will hold (it has $\mathbb{Z}$-basis $B = \langle t, z\rangle\langle y\rangle$, and hence has left $\mathbb{Z}[t]$-basis $B' = B - tB$; and $k(B' - \{1\})$ is a k-bimodule complementing k) and the coproduct is zero.

It would be interesting to know whether there is a similar example where *one* of the given rings is free *both* as a right and as a left k-module, on bases containing 1, and the other has a k-bimodule complement to k.

It would also be interesting to know, given k, *what* k-rings R_1 have faithful and separating coproducts with *all* faithful k-rings R_2?

For some homological results on generalized coproducts of rings, see [22].

9. ANALOGS: SEMIGROUPS, MODULES, ADDITIVE CATEGORIES, ETC.

There are a number of classes of algebraic objects whose definitions are sufficiently similar to those of k-algebras or k-rings that the results of the preceding sections go over to them with only the most minor readjustments in

definitions. I shall run through what I think are the most important and representative, sketching only the differences from the cases considered above.

9.1. *Semigroups*

Let X be a set. By a reduction system for the free semigroup $\langle X\rangle$ we shall mean a subset $S \subseteq \langle X\rangle \times \langle X\rangle$. Given $\sigma = (W_\sigma, f_\sigma) \in S$, and $A, B \in \langle X\rangle$, we define the set-map $r_{A\sigma B}\colon \langle X\rangle \to \langle X\rangle$ to take $AW_\sigma B$ to $Af_\sigma B$ and fix all other elements of $\langle X\rangle$. We let I denote the congruence on $\langle X\rangle$ generated by S, which is generated as an equivalence relation by the set of pairs $(AW_\sigma B, Af_\sigma B)$. Given a semigroup partial ordering "$\leqslant$" on $\langle X\rangle$ compatible with S, for each $C \in \langle X\rangle$ we define I_C to be the equivalence relation generated by those pairs $(AW_\sigma B, Af_\sigma B)$ with $AW_\sigma B \leqslant C$.

It is now obvious how to define ambiguities of S, and the resolvability or resolvability relative to "$\leqslant$" of such ambiguities. The proof of the analog of Theorem 1.2 goes over immediately (without even the need for a Lemma 1.1, whose purpose was to handle certain difficulties with linear combinations of monomials.) For an application see [3, Sect. 7]. One can also get semigroup analogs of Theorem 6.1 and Proposition 7.1. The former, for instance, deals with writing certain extensions R of a semigroup k as unions of sets P_A closed under the right and left actions of k.

One can obtain normal form results in *groups* by treating a group G generated by a set X as a semigroup generated by the set $X \cup X^{-1}$, with appropriate additional relations. For an example, see [53, Sect. 7]. Usually, however, this method is awkward (cf. Sect. 11.1 below) and others are more useful (see Sect. 11.2).

9.2. *Nonunital k-Algebras, k-Rings and Semigroups*

The methods and results of the unital case go over exactly. (A nonunital k-ring means a k-bimodule R with a nonunital ring-structure which is k-bilinear.) Note only that in considering reductions $r_{A\sigma B}$ we must allow A and B to assume the value 1 even though there is no "1" in the objects under consideration.

9.3. *Categories*

A semigroup with 1 is a one-object category, and in general, categories C on a given object-set $\mathscr{P}$ can be presented by generators and relations like semigroups.[1] The generating family X will be a system of sets $(X(q, p))_{q,p\in\mathscr{P}}$, each $X(q, p)$ representing a family of generators introduced in the Hom-set $C(q, p)$. One forms the free category $\langle X\rangle$ on such a system [39, II.7], then introduces relations

[1] Strictly, when speaking of an arbitrary category I should refer to a *class* of objects, generators, relations etc. But I shall be sloppy and say "set" to keep our language as close as possible to that of preceding sections. If C is a category, I shall write $C(q, p)$ for the set of morphisms from p to q in C.

which equate pairs of elements in the sets $\langle X\rangle(q, p)$. In particular, we may take a set of relations $S \subseteq \bigcup_{q,p}(\langle X\rangle(q, p))^2$ which we wish to use to reduce elements to a canonical form. After introducing a partial ordering "$\leqslant$" with descending chain condition on $\langle X\rangle$ (that is, on $\bigcup \langle X\rangle(q, p)$) one again gets the result: If all ambiguities of S are resolvable relative to $\leqslant$ then the set $\langle X\rangle_{\text{irr}}$ of words irreducible under S forms a transversal to the congruence I generated by S, and thus gives a canonical form for elements (morphisms) of the category (with object-set $\mathscr{P}$) $C = \langle X\rangle/I$.

As an application, we note that given the above result, the two pages of argument establishing the normal form result of [64, Theorem 24.1] reduce to the trivial observation that the reduction system introduced there has *no* ambiguities!

9.4. *k-Linear Categories*

If k is a commutative ring, then a *k-linear category* is a category C with a k-module structure on the sets $C(q, p)$, such that the composition-maps $C(r, q) \times C(q, p) \to C(r, p)$ are k-bilinear. These are "k-algebras with several objects" (cf. [64]). Reduction systems for k-linear categories with object-set $\mathscr{P}$ can now be defined in the obvious manner. Sets of generators and reductions are partitioned by $\mathscr{P} \times \mathscr{P}$ as in the preceding case, and k-linearity is treated as in Section 1.

9.5. *Modules, Bimodules and Tensor Products*

Consider a k-linear category C with just two objects, p and q, and such that only $C(p, p)$ and $C(q, p)$ are nontrivial; i.e. $C(q, q) = k$, $C(p, q) = \{0\}$. Such a category is described by giving one k-algebra $R = C(p, p)$ and one right R-module $M = C(q, p)$. A presentation of such a C by generators and a reduction system is of course a special case of the general formulation indicated in the preceding paragraph. If we look at what that comes to in this case, we see that it breaks down into two parts: first, a generating set $X_R = X(p, p)$ and a reduction system S_R for the k-algebra R, exactly as in Section 1, and then a generating set $X_M = X(q, p)$ and reduction system S_M for M as a right R-module. Here S_M consists of pairs (W_σ, f_σ) with $W_\sigma \in \langle X\rangle(q, p) = X_M\langle X_R\rangle$, the free right $\langle X_R\rangle$-set on X_M, and $f_\sigma \in k\langle X\rangle(q, p)$ which is similarly the free right $k\langle X_R\rangle$-module on the basis X_M. After resolving the ambiguities involving S_R alone, we are left with those relating to the module structure of M. Here the overlap ambiguities (the more important sort; cf. 5.1 above) will have the form (σ, τ, xA, B, D), where $\sigma = (xAB, f_\sigma) \in S_M$ and $\tau = (BD, f_\tau) \in S_R$ $(x \in X_M, A, B, D \in \langle X_R\rangle)$. If all ambiguities are resolvable relative to an appropriate partial ordering, then we get a k-basis $\langle X\rangle(q, p)_{\text{irr}}$ for the R-module M. There should be simple and interesting applications, but I do not know any.

If we modify the above considerations by only assuming $C(p, q)$ trivial but

not $C(q, q)$, then the category C is described by two k-algebras $R = C(p, p)$, $R' = C(q, q)$, and an (R', R)-bimodule$_k$, $B = C(q, p)$. (The subscript k means that the actions of the commutative ring k on the right and left are assumed to agree.) So we also have a Diamond Lemma for such bimodules.

Consider, finally, a k-linear category C with three objects, p, q and r, presented by generators and relations in $C(r, q)$, $C(q, q)$ and $C(q, p)$. Then $C(q, q)$ will be a k-algebra R, $C(r, q)$ a right R-module M, and $C(q, p)$ a left R-module N. There will in general be one more nontrivial Hom-module: $C(r, p)$. Since no new generators or relations were assumed to be introduced in this module, it will take the form $M \otimes_R N$. Now if one has reduction systems for R, M and N, for which one has verified that all ambiguities are resolvable, one can then look at the one remaining sort of ambiguities for our system, those corresponding to ambiguously reducible monomials of $\langle X\rangle(r, p)$, which will have the form $(\sigma, \tau, xA, B, Cy)$ with $\sigma \in S_M$, $\tau \in S_N$. Any new equations obtainable from these will represent the "interaction" between the module structures of M and N which occurs when one forms the tensor product $M \otimes_R N$. Hence such interaction can be studied by examining ambiguities. In this situation one may also, of course, introduce generators and relations in $C(r, r)$ and/or $C(p, p)$, making M and/or N bimodules, with these algebras acting on the other side.

9.6. *Extensions of Ab-Categories*

Following [39] we shall call a $\mathbb{Z}$-linear category an "*Ab*-category." These are "rings with several objects" [64]. If k is an *Ab*-category (note the change in notation from the preceding section!) the proper definition of a k-bimodule is a $\mathbb{Z}$-bilinear functor from $k \times k^{op}$ to the category of Abelian groups. Starting from this definition, one can get very close formal analogs of Theorem 6.1 and Proposition 7.1; but I will leave this to the specialist. I will sketch instead an *Ab*-category version of Theorem 6.1 which is less general than this, but easier to picture. (It is, in fact, the case of the general result in which the *Ab*-category k is "totally disconnected"—$k(q, p) = 0$ for $q \neq p$—and each of the given generating bimodules M_x "lives" at only one pair $(p, q) \in Ob(k \times k^{op})$.)

Let $\mathscr{P}$ be a set, and for each $p \in \mathscr{P}$, let there be given a ring $k(p)$. Let $X = (X(q, p))_{q,p\in\mathscr{P}}$ be a family of sets, and $\langle X\rangle$ the corresponding free category with object set $\mathscr{P}$, as in 9.4. For each $x \in X(q, p)$ $(q, p \in \mathscr{P})$ let us be given a $(k(q), k(p))$-bimodule M_x, and for each $A = x_1 \cdots x_n \in \langle X\rangle(q, p)$ let P_A denote the $(k(q), k(p))$-bimodule $M_{x_1} \otimes \cdots \otimes M_{x_n}$ (each "$\otimes$" being taken over the appropriate $k(r)$. As a degenerate case we taken $P_{1_p} = k(p)$ $(p \in \mathscr{P})$.) Then we can define an *Ab*-category $k\langle M\rangle$ with object-set $\mathscr{P}$ by taking $k\langle M\rangle(q, p) = \bigoplus_{A\in\langle X\rangle(q,p)} P_A$, and making composition in $k\langle M\rangle$ correspond to tensor multiplication, just as for the k-ring $k\langle M\rangle$ of Section 6. It is now easy to see how to extend the bimodule-map version of a reduction system described in Section 6 to this many-object setting, and get the analog of Theorem 6.1.

9.7. *Truncated Filtered Rings*

Truncated filtered rings were defined and used in [20, pp. 81–83]. In the paragraph that begins at the bottom of p. 82 thereof, one realizes that some form of the Diamond Lemma for these objects is implicitly being called upon. (This was more explicit in the section of my thesis on which that passage was based.) I shall give below some observations from which the reader interested in these objects who has carefully read [20, pp. 81–82] can work out the details of the required form of the Diamond Lemma (an analog of Proposition 7.1), and justify the arguments made there. Other readers should skip this section.

In presenting a truncated filtered ring of height h, one introduces generators of specified degrees $\leqslant h$, and relations among these. If the desired ring is to have (say) the property that a certain generator x of degree i and a generator y of degree i' have product xy of degree $j < i + i'$, one may achieve this by including in the presentation another generator z of degree j, and a reduction (xy, z). Thus, more generally, the type of normal forms one looks for are those in which the degree of an element of R is the formal degree of its normal form. The expressions for elements of R are to be all irreducible expressions of formal degree $\leqslant h$ in the given generators. In verifying that one has such a normal form, one need only check that ambiguities of formal degree $\leqslant h$ can be resolved.

Now what is asserted in [20, paragraph beginning at the bottom of p. 82] is that *certain* reduction systems which lead to canonical forms for truncated filtered rings of height h will also lead to canonical forms when considered as presenting truncated filtered rings of height $h + 1$. For an arbitrary reduction system this might fail, because there are more ambiguities of height $\leqslant h + 1$ to be checked than of height $\leqslant h$. (A consequence is that the universal construction pushing a truncated filtered ring of height h into one of height $h + 1$ is not always injective [20, p. 86, Exercise 2.].) But for the particular well-behaved class of objects being considered at that point, one gets presentations in which all relations are of the form (19) and none of the form (20) (Sect. 7 above), and as a result, there are no ambiguities of degree higher than the highest degree of a generator, which is $\leqslant h$.

10. Contrasts: Other Sorts of Algebraic Systems

In this section the word *algebra* will be used in the sense of universal algebra: An *algebra-type*, T, will mean a set of symbols (intended to denote operations) with a nonnegative integer (the intended arity) associated to each. An algebra of type T, or T-algebra, will mean a set, given with a family of operations, one for each symbol in T, of the prescribed arities. A variety V of algebras of type T will mean the subclass of the algebras of type T determined by some family of algebras identities; equivalently, by Birkhoff's Theorem [20, Theorem IV.3.1], a class of algebras of type T closed under direct products, subalgebras, and homomorphic images.

10.1. *Canonical Forms in Free Algebras*

A prerequisite for getting a result like Theorem 1.2 for algebras in an arbitrary variety V would seem to be that a canonical form should be known in the *free* algebras of V, in terms of which one can calculate easily.

The very general problem of establishing a canonical form for free algebras, given a presentation of the variety by a list of operations and identities, is treated by Knuth and Bendix [34]. Their approach is, in fact, parallel to that of this paper. They seek a system of *identities* with which to reduce an arbitrary expression in a free algebra. To test for (what we have called) reduction-uniqueness, they show that it suffices to check minimal test-cases, analogous to our ambiguities. If one of these is not resolvable, it is used to get a new identity, as in Section 2 above.

The parallelism of the two cases is not an accident. Lawvere [36] (cf. [38, 6]) has shown that the operations and identities of a variety V can be looked at as the morphisms and relations holding in a category θ whose object-set has the form $\{0, 1, 2, ...\}$, where "n" is the n-fold coproduct of "1". From this point of view, the ambiguously reducible expressions in Knuth and Bendix's development are essentially overlap ambiguities in the category θ. (But this does not reduce their situation to a subcase of that of Section 9.3 above, because the condition that n be the n-fold coproduct of 1 puts conditions on the morphisms of θ outside of the sort we considered; see below. The one case where this effect can be ignored is when all the operations of V are *unary*. Such varieties V correspond precisely to semigroups with 1.)

A severe limitation of the method of [34] is exemplified by its inability to handle the commutative identity, $a \cdot b = b \cdot a$ [34, Example 18]. To introduce a reduction $a \cdot b \mapsto b \cdot a$ would yield, for all expressions x and y, the reduction $x \cdot y \mapsto y \cdot x$ and also $y \cdot x \mapsto x \cdot y$; clearly this gives a nonterminating reduction-procedure. But it appears likely that the machinery of [34] can be refined to handle such cases. For instance, if one introduces a total ordering on all expressions, including the variables, then commutativity could be expressed by a reduction rule "$a \cdot b \mapsto b \cdot a$ when $a > b$."

For some other work on word problems in free algebras see [46, 68].

10.2. *k-Linear T-Algebras*

In this section we will introduce some useful language. Let T be an algebra-type, and k a commutative ring. Then by a "k-linear T-algebra" we will mean a k-module given with a structure of T-algebra such that all operations are k-multilinear. If A is a k-linear T-algebra, let $U(A)$ denote its underlying T-algebra, and if B is a T-algebra, let kB denote the k-linear T-algebra made from the free k-module on B by extending the T-operations of B multilinearly.

By a *regular* T-identity let us mean an equation $f = g$, where f and g are each expressions in the operations of T and some indeterminates, such that each

indeterminate occurs exactly once in f and once in g. Familiar examples of regular identities are $x \cdot (y \cdot z) = (x \cdot y) \cdot z$, $x \cdot e = x$, $x \cdot y = y \cdot x$, and $(x \cdot y)^{-1} = y^{-1} \cdot x^{-1}$. Examples of nonregular identities are $x \cdot x^{-1} = e$, and any identity of a *k-linear* T-algebra that involves the linear structure rather than just the T-structure, such as $x \cdot y = -y \cdot x$.

If I is any set of regular T-identities, then there is a very close connection between the variety V_I of T-algebras satisfying I, and the variety $V_{k,I}$ of k-linear T-algebras satisfying I: It is easy to see that a k-linear T-algebra A will lie in $V_{k,I}$ if and only if $U(A)$ lies in V_I; less trivially, a T-algebra B will lie in V_I if and only if kB lies in $V_{k,I}$. Hence we get a one-to-one correspondence $V_I \leftrightarrow V_{k,I}$ between varieties of T-algebras defined by sets of regular identities and varieties of k-linear T-algebras defined by sets of regular identities. We shall call such V_I and $V_{k,I}$ "corresponding varieties." Thus, if T is the type with just one binary operation, "$\cdot$", then the variety of all T-algebras (called "groupoids," or "magmas"), the variety of all semigroups, and the variety of all commutative semigroups correspond respectively to the k-linear varieties of all k-algebras, all associative k-algebras, and all commutative associative k-algebras. (And if we add to T one zeroary operation e, the above correspondences hold with "unital" thrown in.) On the other hand, the variety of groups (with operations $\cdot$, e, $(\ )^{-1}$) corresponds to no k-linear variety, and the k-linear variety of Lie algebras corresponds to no non-linear variety, because neither can be defined by regular identities.

If V_I and $V_{k,I}$ are corresponding varieties, note that a normal form for free algebras of V_I immediately yields a normal form for free algebras of $V_{k,I}$.

Returning to *arbitrary* varieties of k-linear T-algebras, I would guess that the methods of Knuth and Bendix discussed in the preceding section should extend nicely to these—not by naively introducing the k-module structure and the k-multilinearity of the T-operations on a par with the other operations and identities, but by working with k-linear combinations of T-expressions, k-linear reductions, etc., in the spirit of preceding sections of this paper.

10.3. *Commutative Algebras*

Let us consider how to formulate a version of Theorem 1.2 for commutative associative k-algebras. If X is a set, $[X]$ will denote the free commutative semigroup on X with 1, and $k[X]$ the usual polynomial algebra on X. Now the proof of Theorem 1.2 was based on the property of the free semigroup $\langle X \rangle$, that if two elements W_σ and W_τ both occurred as subwords of some element D, then either these occurrences were disjoint, or D contained an occurrence of some word formed by overlapping the two given words. In the commutative situation, the analog of "A is a subword of B" is "B is a multiple of A in $[X]$"; reductions can be written $r_{A\sigma}$, this map taking AW_σ to Af_σ; and we see that an element D will be a multiple of both W_σ and W_τ if and only if it is a multiple of $W_\sigma \vee W_\tau$, their l.c.m. in $[X]$. Hence in place of finding and resolving all ambiguities, we

must establish, for each $\sigma, \tau \in S$, that the results of the two ways of reducing $W_\sigma \vee W_\tau$ can be reduced to a common value. With this modification, Theorem 1.2 goes over to commutative algebras.

Now consider the analog of the considerations of Section 5.3. Say k is a field and X finite, and that we have chosen a semigroup well-ordering of $[X]$, and are given a finite set of relations to impose on $k[X]$. As in that section, we write these as formulae for reducing the maximal monomial of each relation. For each σ, τ in the resulting reduction system, we compare the two ways of reducing $W_\sigma \vee W_\tau$, and if they disagree, this gives us a new reduction equation.

Now note that every time the above process introduces a new reduction ρ into S, the term W_ρ will *not* be a multiple of any W_σ for σ already in S. This means that the ideal of $k[X]$ generated by the elements W_σ will increase at every such step. But $k[X]$ is Noetherian. Hence only a finite number of new elements can be introduced before the process terminates, and gives us our normal form—in contrast to the situation for reduction systems on $k\langle X\rangle$. In particular, the word problem for finitely presented commutative associative k-algebras is solvable. This has been known for a long time [29] (cf. [51]).

Of course, the same results will hold in the corresponding non-k-linear variety, that of Abelian semigroups.

10.4. *Other Varieties with Solvable Word Problems*

Evans [25] has shown that the word problem is solvable for finitely presented algebras in the variety having *one* operation f_i of *each* positive arity i, and only certain commutativity-like identities: namely, for each i, a subgroup G_i of the permutation group S_i is assumed given, and f_i is assumed invariant under G_i-permutations of variables. It appears that the assumption "exactly one operation of each arity" is irrelevant; one only need assume given a set of operations f of prescribed arities $n(f)$, and for each f, a subgroup $G_f \subseteq S_{n(f)}$.

Evans' commutativity-like identities are regular in the sense of Section 10.2, and it seems that his results should extend to the corresponding k-linear category, for k a field. In this k-linear situation, a natural generalization of a subgroup $G_f \subseteq S_{n(f)}$ is a left ideal C_f in the group algebra $kS_{n(f)}$; one would associate to each element $a = \sum_{\pi \in S_{(n)_f}} \alpha_\pi \pi \in C_f$ the identity $\sum_{\pi \in S_{(n)_f}} \alpha_\pi f(x_{\pi(1)}, \ldots, x_{\pi(n)}) = 0$ in our algebras. I suspect that Evans' results should generalize to varieties defined by this sort of identities.

(Caveat: If $\cdot$ is a binary operation, then identities such as $(x \cdot y) \cdot z = (z \cdot x) \cdot y = (y \cdot z) \cdot x$ are *not* covered by Evans' result, since $((\cdot)\cdot)$ is not a primitive operation. Precisely the fact that, unlike the associative law, his identities do not "break through" parentheses appears to be the reason things are so simple in this case.)

Note that, taking all $G_f = \{e\}$, Evans' results apply to varieties with *no* identities. In fact, for such varieties one can easily get a version of Theorem 1.2.

Here there are only inclusion ambiguities, and these can be eliminated as in Section 5.1, to give reduction systems without ambiguities.

Gluhov [27] claims to obtain normal forms in all finitely presented *lattices*, and to show that in fact every finitely presented lattice has a distinguished "minimal" set of generators and relations (but the completeness of his proof has been questioned). Evans [24, 26] obtains similar results for *loops* and some similar varieties, except that the uniqueness of the minimal presentation is modulo a specified group of automorphisms. For some other work on normal forms in lattices, cf. [21, 23, 59] and items cited in [66, esp. p. 110].

Normal forms for finitely presented (=finite) *distributive lattices* and *Boolean algebras* are easily obtained.

For some other results and problems see [67].

The results cited here are those that have been brought to my attention; many others may be known.

10.5. *Some Open Cases*

For k a field, there are a great number of proper subvarieties of the variety of associative unital k-algebras. The most important of these are, for each n, the variety $\mathscr{M}(n, k)$ generated by all $n \times n$ matrix algebras over commutative k-algebras. (These are, except when k is finite, the only varieties generated by prime k-algebras.) The free algebra on r indeterminates in $\mathscr{M}(n, k)$ can be described as the k-algebra generated by matrices with distinct commutative-indeterminate entries, i.e., a subalgebra of the matrix ring $M_n(k[x_{qij} \mid q \leqslant r; i, j \leqslant n])$. As one can compute explicitly in this matrix ring, the word problem for these free algebras is solvable. But very basic questions are open: It is not known whether these varieties are determined by finitely many identities, except that Razmyslov [48] has proved this for $n = 2$. Neither explicit k-bases for the free algebras, nor even the dimension of the space spanned by homogeneous multilinear monomials of degree r, as a function of r, are known for these varieties, nor for the varieties determined by particular single identities, except for some very special ones. Regev and others [49, 72] have been working on these problems. A general analog of Theorem 1.2 for algebras in these varieties still seems far out of reach. The theory of "algebras with polynomial identities" has a large literature; a few general references and further works relevant to the study of the identities themselves are [2, 9, 50, 62, 63, 65].

A normal form is known for elements of free Lie algebras. For a nicely motivated development, see [28] and for extensive further work [57]. Note also [12, 76]. So this variety seems a good candidate in which to try for an analog to Theorem 1.2. It should also be possible to solve word problems in some Lie algebras $\mathbf{L}$ by applying Theorem 1.2 to the associative algebras $k[\mathbf{L}]$. Some related varieties are those of Jordan algebras [32] and of "graded Lie algebras" with identities twisted by the grading [42, 73]. Theorems 1(ii) and 3 of [13] suggest

that the k-linear variety with one binary operation and one quartenary operation, subject to just those identities holding for the operations $(x, y) = \frac{1}{2}(xy + yx)$, $(w, x, y, z) = \frac{1}{2}(wxyz + zyxw)$ in associative algebras may have good properties. (See [45] for a general survey of varieties of k-linear algebras with one binary operation.)

Finding a normal form for free modular lattices is an outstanding open problem of lattice theory [61]. I would suppose the same is true for the smaller variety generated by all lattices of submodules of modules. The modular identity is the simplest nontrivial lattice identity that these satisfy, but they also satisfy identities not following from this one [33; 69, 13.6]. For a negative result on word problems in modular lattices, with an unexpected application to diagram-chasing theorems in additive categories see [70].

One might look for normal forms in "involution algebras" [6]. Perhaps, to avoid confusion with k-algebras with involution, one should rename these "involution systems." An interesting related variety is that of "conjugoids," defined by Alan G. Waterman as sets with a binary operation satisfying $x \cdot x = x$ and $x \cdot (y \cdot z) = (x \cdot y) \cdot (x \cdot z)$ (personal communication).

10.6. *Classes of Algebras Other Than Varieties*

Let me simply mention two important cases.

A *quasivariety* [16, 40], is a class of algebras of a given type T defined by a family of *Horn sentences* (or *conditional identities*):

$$\forall x_1, \ldots, x_r (f_1 = g_1) \wedge \cdots \wedge (f_n = g_n) \Rightarrow (f_{n+1} = g_{n+1}) \qquad (n \geqslant 0), \quad (24)$$

where f_i, g_i are expressions in the x_j.

The classes of torsion-free groups, cancellation semigroups, and rings without nonzero nilpotent elements are examples. Rings without zero-divisors or without idempotent elements $\neq 0, 1$ are not. Some classes of algebras which can be shown by another criterion [40, Theorem 5.11.2] to be a quasivariety are the class of rings embeddable in $n \times n$ matrix rings over commutative rings, and the class of groups whose group algebras over a specified domain have no nontrivial nilpotent elements.

Free algebras in quasivarieties are always the same as the free algebras in the varieties they generate; but if we want to establish a normal form for elements of an algebra presented by finitely many generators, relations and Horn sentences, a basic problem is that of finding the solutions to equations on the left-hand side of (24), so that one can then apply the right-hand side.

Even more basic problems are presented by the class of *division rings*. Here one does not even know ab initio which expressions in a set of generators will make sense; for to say whether f^{-1} is defined one must *first* decide whether $f = 0$. As a consequence, there are no "free division rings" in the standard sense. However, it has recently been found that if instead of looking at *homomorphisms* among division rings (a very limited class of maps because they must all be

embeddings) one considers *specialization* maps, then "free division rings" and certain "division rings presented by generators and relations" will exist, and the study of the structures of these has been begun. See [7, 77, 78] and references cited [7, Sect. 12].

Projective planes resemble division rings in that the join of two points p and q, or the point of intersection of two lines p and q, is defined only if $p \neq q$. There is the difference, however, that whereas in a division ring the equation $ax = b$ has *no* solution if $a = 0, b \neq 0$, in a projective plane the problem "find a line through points p and q" has solutions, but simply not a unique one, if $p = q$. This makes it easy to get projective planes with a weakened version of the conventional sense of "freeness." See [30, Sect. XI]. I do not know whether the analog of specilizations has been studied for projective planes; it is easy to define.

11. General Observations on Canonical Form Results

The three subsections below are independent of one another.

11.1. "*The Stumbling-Block*"

The following example illustrates a basic stumbling-block in formulating and proving normal-form results for many sorts of algebras. Consider a group G presented by generators $x_1, \ldots, x_n$ and some set of relations. Suppose we set up a system of reduction rules on elements of the *free group* F on $x_1, \ldots, x_n$, with the understanding that

> whenever $A \mapsto A'$ is one of our rules, then $f(x_1, \ldots, x_n, A) \mapsto f(x_1, \ldots, x_n, A')$ is an allowable reduction, for any expression f in $n+1$ variables. (25)

Then for every $B \in F$, and any one of our rules $A \mapsto A'$, we note that $B = AA^{-1}B$, which can be reduced by (25) to $A'A^{-1}B$. Hence no element of F is irreducible under our system! Thus no reduction procedure satisfying (25) can give canonical forms for groups; and the same applies to rings in view of their additive group structure.[2]

The way we got around that difficulty in this paper was to apply (25) only to

[2] So, for instance, Theorem III.9.3 of Cohn [16], often quoted as a reference in normal form arguments in ring theory, cannot in fact be so applied, if a "direct move" is defined as in the paragraph preceding that theorem (which is essentially (25)) since the termination condition (i) of the hypothesis of that theorem can then never be satisfied. Cohn tells me that the indicated paragraph should only be read as a suggestion of how "direct move" might be defined, and agrees that something like Lemma 1.1 of this paper is needed to rigorously justify the application of that Theorem to rings. Presumably one should take for one's "direct moves" what we have called "reductions."

monomials f, and then to extend our reductions to $k\langle X\rangle$ k-linearly. Thus our maps $r_{A\sigma B}$ act unselectively on all occurrences of the monomial $AW_\sigma B$, so that one *cannot*, for instance, write

$$0 = AW_\sigma B - AW_\sigma B \underset{r_{A\sigma B}}{\mapsto} Af_\sigma B - AW_\sigma B. \tag{26}$$

This nonselectivity led to the problem: If we know, say, that $f_\sigma C$ and Af_τ can be reduced to a common value, it is not clear that $f_\sigma C + d$ and $Af_\tau + d$ can also be so reduced, because the reductions one applies to the two expressions might affect d differently. This is what made the proof of Theorem 1.2 nontrivial, requiring Lemma 1.1 and the careful inductive use of reduction-uniqueness.

In the end we found that we can have our cake and eat it too, if we are careful: Theorem 1.2(a′) allows us to "use (26)," that is, to add terms $AW_\sigma B - Af_\sigma B$ to an expression we are studying in verifying the resolvability of an ambiguity, *if* $AW_\sigma B$ is $<$ the original word in question. Lemma 4.1 similarly allows us to apply reductions "selectively" under certain conditions.

Mauldon [41], working on the same problem we treated in Section 2 (a fruitful problem) got around the basic stumbling-block in another way. Where we have used $k\langle X\rangle$, he uses the semigroup-*semiring* $\mathbb{N}(k \times \langle X\rangle)$, where $\mathbb{N}$ is the semiring of nonnegative integers. In this semiring there are no additive inverses, so (25) or an appropriate variant can be used, and many of the difficulties of our Section 1 are avoided. He includes, among his reduction-rules, reductions that "restore" the k-linearity in the final object.

If one abstracts Mauldon's argument, it leads to an analog of Theorem 1.2 in which the ambiguities that must be resolved appear in $\mathbb{N}(k \times \langle X\rangle)$. This is easier to prove than our Theorem 1.2, but the extra work comes in again if one wants to bring it to a fomulation convenient to ring-theorists rather than semiring-theorists.

We note that in Section 9.1 where we indicated that normal forms in groups could be studied by presenting them as semigroups, we were taking an analog of this approach of Mauldon's. We could not follow the approach of Section 1 because the inverse operation in groups does not occur as part of a nice linear structure. But perhaps group theorists can find simplifications appropriate to this case.

Returning to rings, note that even by assuming (25) as we have for *monomials* f, we have restricted the kinds of canonical form we are considering. As an example of a canonical form thus excluded, suppose we wish to present the $n \times n$ matrix ring $M_n(k)$ in terms of the generators $p_i = e_{i1}$ and $q_i = e_{1i}$, writing every element in terms of the basis of elements $p_i q_j$. Then we must have the "reduction" $p_i \mapsto p_i q_1$. But the formalism of this paper would then give the further "reductions" $p_i q_1 \mapsto p_i q_1 q_1 \mapsto \cdots$, which we do not want. (This example is, of course, contrived. Its purpose is to alert us to the fact that we may at some time find it desirable to study normal bases not of the sort given by the formalism

of Section 1. For another sort of nonstandard normal form, see [74, Sect. 2], which assumes only the first paragraph of Section 1 of that paper.)

11.2. *The Other Way of Establishing Normal Form Results*

In proving a normal form result for the elements of an algebra R (in the general sense) defined by a universal condition—freeness, presentation by generators and relations, etc.—it is usually easy to show that every element of R can be expressed in the desired form; the hard part is to show the uniqueness of the reduced expression; that is, that distinct reduced expressions represent distinct elements. There are, in general, two ways of doing this:

> Examine the universal construction of R, and analyse what expressions can fall together therein. (27)

> Construct a model R' satisfying the conditions for which R is universal (though not perhaps the universality itself) and show that distinct reduced expressions represent distinct elements of R'. (28)
> Then by the universality of R, the corresponding elements of R must also be distinct.

Ultimately (27) and (28) are the same (or (27) is a case of (28)), since when one abstractly constructs the universal object R (say, as a set of symbols modulo an equivalence relation) one has to show that it *is* a model of the given conditions. But in practice, they represent quite different techniques. The approach of this paper has been (27). A few words now on (28):

A special version of (28) which is useful for classes of algebras that arise "naturally" is

> Construct the model R' as an algebra of operations etc. on some other object E. (28′)

For example, to get an associative k-algebra, take an algebra of endomorphisms of a k-module; to get a group or semigroup, use a group of permutations, or a semigroup of endomaps, of a set; to get a lattice, use the lattice of closed sets in some closure system ([16]; e.g. the lattice of subalgebras of some algebra). An unexpectedly powerful subcase of this, in turn, is

> Take for this E the set of expressions which you wish to prove is a normal form for R! (Or some set formed therefrom.) (28″)

This trick was introduced by van der Waerden [58], who used it to establish the normal form for a coproduct ("free product") of a family of groups, $\coprod_I G_i$ [35, Vol. II, Sect. 35]. One forms the set E of all reduced words in the "alphabet" $\bigcup_I G_i$, describes a *left action of each* G_i on E, and considers the group G of

permutations of E generated by all these actions. One would expect this approach to lead in a circle, and not have any advantage over the standard verification that the "natural" multiplication on reduced words is a group structure. But in fact, this method cuts away the whole tedious proof of associativity, since composition of permutations is automatically an associative operation; and on the other hand, it is easy to see that if u and v are distinct reduced expressions, then the corresponding products-of-actions-on-E are distinct, since they take the empty word 1 to u and v respectively.

The same method works for coproducts of semigroups, and of k-rings with good k-module structures. See [4] for a proof of Corollary 8.2 without the help of the Diamond Lemma.

In general, (28″) seems to make a satisfactory and elegant alternative to Theorem 1.2 for rings in cases where the reduction procedure is such that when one multiplies a reduced word on the left by an arbitrary letter, the resulting reductions cannot "propagate" to the right, so that the actions of those generators on the free k-module spanned by the reduced words can be concisely described. But in cases like those of Sections 2 and 3, there is no control on such "propagation," and Theorem 1.2 seems the best approach. (The method of (28″) *is* used to prove the Poincaré–Birkhoff–Witt theorem in [52], but this requires a complicated induction, and does not yield nearly as easy a proof as in Section 3 above.)

An exposition of universal constructions and canonical form results at the level of an elementary graduate course, including the ideas of (27)–(28″) is given in [8].

11.3. *Direct Limit Canonical Forms*

Suppose that $\mathfrak{D}$ is a directed partially ordered set, that for every $d \in \mathfrak{D}$, F_d is a family of expressions for some elements of an algebraic system R, on which expressions we are given some sort of reduction procedure, and that for every pair $d < d'$ a map $f_{d'd}: F_d \to F_{d'}$ is given which respects reductions, such that for $d < d' < d''$ and $A \in F_d$, one has $f_{d''d}(A) = f_{d''d'}(f_{d'd}(A))$. Then if every element of R is represented by an expression in *some* F_d, and every expression in every F_d has reduction-unique images in all $F_{d'}$ with d' sufficiently large; and if two expressions represent the same element if and only if their reduced forms eventually agree, then we could say that we have a *direct limit canonical form* for the elements of R.

(This definition is only a rough suggestion, which I leave to other investigators to refine. Perhaps the maps $f_{d'd}$ could do the job of reductions as well. Perhaps $\mathfrak{D}$ should be a more general category than a partially ordered set.)

One very simple example of this is the familiar construction of a localization $C[S^{-1}]$ of a commutative ring C with respect to a multiplicative semigroup $S \subseteq C$. For each $s \in S$ we have a set of formal expressions $F_s = \{as^{-1} \mid a \in C\}$. Whenever $s, st \in S$ we map F_s to F_{st} by $as^{-1} \to (at)(st)^{-1}$. Though we do not

have here a normal form in any strict sense, what we have serves much of the function of a normal form: it gives us a strong handle on computations with elements in the object in question.

What actually led me to the idea of direct limit canonical forms was the annoying observation that Corollary 8.2 above does not give the full result of Cohn [14, Theorem 4.4]. The latter concerns the right module structure of a coproduct of faithful k-rings, $\coprod R_\lambda$, such that for each λ, the right k-module R_λ/k is *flat*. Now flat modules may be characterized as direct limits of free modules, from which it follows that in the above situation, each right k-module R_λ can be written as a direct limit of free modules $F_{\lambda,d} = k \oplus X_{\lambda,d}k$. If each $F_{\lambda,d}$ could be made a k-ring so that R_λ was their direct limit, we could apply Corollary 8.2 to these rings and go to the limit to get Cohn's result. In general this will not be possible; but as an approximation to such a ring structure, note that each $a \in F_{\lambda,d_1}$, $b \in F_{\lambda,d_2}$ have a "product" in some F_{λ,d_3}. So we may hope to get some sort of direct limit canonical forms for the k-rings R_λ, and then apply methods analogous to the proof of Corollary 8.2 to get the full result of [14]. Applying the same ideas to Corollary 8.1, one should get information on the k-bimodule structures of k-rings R_λ which are direct limits of split bimodule extensions $k \oplus M_{\lambda,d}$.

Whether these particular ideas would work I cannot say, but the idea of a direct limit canonical form seems worth keeping in mind.

References

1. Advanced Problem 5082, proposed by the Junior Research Seminar for High School students of Summer 1962, Lehigh University, *Amer. Math. Monthly* **70** (1963), 335.
2. S. A. Amitsur, Polynomial identities, *Israel J. Math.* **19** (1974), 183–199.
3. G. M. Bergman, The index of a group in a semigroup, *Pacific J. Math.* **36** (1971), 55–62.
4. G. M. Bergman, Modules over coproducts of rings, *Trans. Amer. Math. Soc.* **200** (1974), 1–32.
5. G. M. Bergman, Coproducts, and some universal ring constructions, *Trans. Amer. Math. Soc.* **200** (1974), 33–88.
6. G. M. Bergman, Some category-theoretic ideas in algebra, *in* "Proceedings of the 1974 (Vancouver) I.C.M.," Vol. 1, pp. 285–296.
7. G. M. Bergman, Rational relations and rational identities in division rings, *J. Algebra* **43** (1976), 252–266 and 267–297.
8. G. M. Bergman, General theory of algebraic structures, Course notes, Chaps. 1–3, Berkeley, 1974. (Available from author.)
9. G. M. Bergman, A lemma of Razmyslov, and central polynomials, unpublished.
10. G. M. Bergman and A. P. Doohovskoy, On subsemigroups of free semigroups, and quasivarieties, in preparation.
11. G. Birkhoff, Representability of Lie algebras and Lie groups by matrices, *Ann. of Math.* **38** (1937), 526–532.
12. L. A. Bokut', Unsolvability of the word problem, and subalgebras of finitely presented Lie algebras, *Izv. Akad. Nauk S.S.S.R. Ser. Mat.* **36** (1972), 1173–1219 (Russian).

13. P. M. COHN, Homomorphic images of special Jordan-algebras, *in* "Proceedings of the 1954 (Amsterdam) I.C.M."
14. P. M. COHN, On the free product of associative rings, *Math. Z.* **71** (1959), 380–398. (This work is continued in [15, 18].)
15. P. M. COHN, On the free product of associative rings. II. The case of skew fields, *Math. Z.* **73** (1960), 433–456.
16. P. M. COHN, "Universal Algebra," Harper & Row, New York, 1965.
17. P. M. COHN, Some remarks on the invariant basis property, *Topology* **5** (1966), 215–228.
18. P. M. COHN, On the free product of associative rings, III, *J. Algebra* **8** (1968), 376–383; Erratum, **10** (1968), 123.
19. P. M. COHN, Dependence in rings. II. The dependence number, *Trans. Amer. Math. Soc.* **135** (1969), 267–279.
20. P. M. COHN, "Free Rings and Their Relations," Academic Press, New York, 1971.
21. R. A. DEAN, Completely free lattices generated by partially ordered sets, *Trans. Amer. Math. Soc.* **83** (1956), 238–249.
22. W. DICKS, Mayer–Vietoris presentations over colimits of trees of rings, *Proc. London Math. Soc.* (3) **34** (1977), 557–576.
23. R. P. DILWORTH (Ed.), "Lattice Theory," Proceedings of Symposia in Pure Mathematics, Vol. II, Amer. Math. Soc., Providence, R.I., 1961.
24. T. EVANS, On multiplicative systems defined by generators and relations. I. Normal form theorems, *Proc. Cambridge Philos. Soc.* **47** (1951), 637–649.
25. T. EVANS, A decision problem for transformations of trees, *Canad. J. Math.* **15** (1963), 584–590.
26. T. EVANS, The isomorphism problem for some classes of multiplicative systems, *Trans. Amer. Math. Soc.* **109** (1963), 303–312.
27. M. M. GLUHOV, On the problem of isomorphism of lattices, *Dokl. Akad. Nauk S.S.S.R.* **132** (1960), 254–256; *Soviet Math. Dokl.* 1 (1960), 519–522.
28. P. HALL, Some word problems, *J. London Math. Soc.* **33** (1958), 482–496.
29. G. HERMANN, Die Frage der endlich viele Schritte in der Theorie der Polynomideale, *Math. Ann.* **95** (1926), 31–65.
30. D. R. HUGHES AND F. C. PIPER, "Projective Planes," Springer Graduate Texts in Mathematics, No. 6, Springer, Berlin/New York, 1973.
31. N. JACOBSON, "Lie Algebras," Interscience, New York, 1962.
32. N. JACOBSON, "Structure and Representation of Jordan Algebras," A.M.S. Colloquium Publications, Vol. 39, Amer. Math. Soc., Providence, R.I., 1968.
33. B. JÓNSSON, Modular lattices and Desargues' theorem, *Math. Scand.* **2** (1954), 295–314.
34. D. E. KNUTH AND P. B. BENDIX, Simple word problems in universal algebras, *in* "Computational Problems in Abstract Algebra" (J. Leech, Ed.), pp. 263–297, Pergamon, New York, 1969.
35. A. G. KUROSH, "Theory of Groups" (translated by K. A. Hirsch), Chelsea, New York, 1960.
36. F. W. LAWVERE, "Functional Semantics of Algebraic Theories," Ph.D. Thesis, Columbia University, 1963.
37. W. G. LEAVITT, Modules without invariant basis number, *Proc. Amer. Math. Soc.* **8** 1957), 322–328.
38. F. E. J. LINTON, Some aspects of equational categories, *in* "Proceedings of the Conference on Categorical Algebra, La Jolla, 1965," pp. 88–94, Springer, Berlin/New York, 1966.

39. S. MAC LANE, "Categories for the Working Mathematician," Graduate Texts in Mathematics, No. 5, Springer, Berlin/New York, 1971.
40. A. I. MAL'CEV, "Algebraic Systems" (translated by B. D. Seckler and A. P. Doohovskoy), Grundlagen der mathematischen Wissenschaften, Band 192, Springer, Berlin/New York, 1973.
41. J. G. MAULDON, Nonorthogonal idempotents whose sum is idempotent, *Amer. Math. Monthly* **71** (1964), 963–973.
42. M. W. MILNOR AND J. C. MOORE, On the structure of Hopf algebras, *Ann. of Math.* **81** (1965), 211–264.
43. M. H. A. NEWMAN, On theories with a combinatorial definition of "equivalence," *Ann. of Math.* **43** (1942), 223–243.
44. O. ORE, Theory of noncommutative polynomials, *Ann. of Math.* **34** (1933), 480–508.
45. J. M. OSBORNE, Varieties of algebras, *Advances in Math.* **8** (1972), 163–369.
46. D. PIGOZZI, Universal equational theories and varieties of algebra, to appear.
47. H. POINCARÉ, Sur les groupes continus, *Trans. Cambridge Philos. Soc.* **18** (1900), 220–255.
48. YU. P. RAZMYSLOV, The finite generability of the identities of a matrix algebra of order 2 over a field of characteristic 0, *Algebra i Logika* **12** (1973), 83–113.
49. A. REGEV, The T-ideal generated by the standard identity ${}_3[x_1, x_2, x_3]$, *Israel J. Math.* **26** (1977).
50. S. ROSSET, A new proof of the Amitsur-Levitski identity, *Israel J. Math.* **23** (1975), 187–188.
51. A. SEIDENBERG, On the length of a Hilbert ascending chain, *Proc. Amer. Math. Soc.* **29** (1971), 443–450.
52. J.-P. SERRE, "Lie Algebras and Lie Groups," Benjamin, New York, 1965.
53. P. B. SHAY, Discoherently associative bifunctors on groups, *in* "Reports of the Midwest Category Seminar V" (J. W. Gray and S. MacLane, Eds.), Springer Lecture Notes No. 195, Springer, Berlin/New York, 1971.
54. J. R. STALLINGS, Whitehead torsion of free products, *Ann. of Math.* **82** (1965), 354–363.
55. E. J. TAFT AND R. L. WILSON, Hopf algebras with nonsemisimple antipode, *Proc. Amer. Math. Soc.* **49** (1975), 269–276.
56. D. E. RADFORD, Operators on Hopf algebras, *Amer. J. Math.* **99** (1977), 139–158.
57. G. VIENNOT, "Algèbres de Lie libres et monoïdes libres," thèse de doctorat d'état, Université Paris VII, 1974.
58. B. L. VAN DER WAERDEN, Free products of groups, *Amer. J. Math.* **70** (1948), 527–528.
59. P. WHITMAN, Free lattices, *Ann. of Math.* **42** (1941), 325–330; **43** (1942), 104–115.
60. E. WITT, Treue Darstellung Liescher Ringe, *J. Reine Angew. Math.* **117** (1937), 152–160.
61. G. BIRKHOFF, "Lattice Theory," 3rd ed., A.M.S. Colloquium Publication No. 25, Amer. Math. Soc., Providence, R.I., 1967.
62. N. JACOBSON, "PI-Algebras, An Introduction," Springer Lecture Notes in Mathematics No. 441, Springer, Berlin/New York, 1975.
63. B. KOSTANT, A theorem of Frobenius, a theorem of Amitsur–Levitzky, and cohomology theory, *J. Math. Mech.* (now *Indiana Univ. J. Math.*) **7** (1958), 237–264.
64. B. MITCHELL, Rings with several objects, *Advances in Math.* **8** (1972), 1–161.
65. C. PROCESI, The invariant theory of $n \times n$ matrices, *Advances in Math.* **19** (1976), 306–381.
66. M. M. GLUHOV, I. V. STELLETSKI, AND T. S. FOFANOVA, Lattice Theory (Russian, survey article), *in* "Algebra, Topologiya, Geometriya 1968," Akad. Nauk SSSR,

Moscow, 1970. (Note: The proof in item [8] of the bibliography of [66] appears to be defective, as is also the proof in item [5] of the bibliography of *that* paper.)
67. T. Evans, Some solvable word problems, *Proceedings* of a conference on decision problems in algebra held in Oxford, July 1976. To appear, North-Holland.
68. D. Pigozzi, Base-undecidable properties of universal varieties, *Algebra Universalis* **6** (1976), 193–223.
69. P. Crawley and R. P. Dilworth, "Algebraic Theory of Lattices," Prentice–Hall, Englewood Cliffs, N.J., 1973.
70. G. Hutchinson, Recursively undecidable word problems of modular lattices and diagram chasing, *J. Algebra* **26** (1973), 385–399. For further work along these lines see next item.
71. G. Hutchinson, Embedding and unsolvability theorems for modular lattices, *Algebra Universalis* **7** (1977), 47–84.
72. A. Regev, The representations of S_n and explicit identities for P.I. algebras, *J. Algebra* **51** (1978), 25–40.
73. L. Corwin, Y. Ne'eman, and S. Sternberg, Graded Lie algebras in mathematics and physics, *Rev. Modern Physics* **47** (1975), 573–603.
74. G. M. Bergman, The global dimension of mixed coproduct/tensor-product algebras, to appear.
75. G. M. Bergman, Embedding filtered rings in completed graded rings, to appear.
76. G. M. Bergman, The Lie algebra of vector fields on $\mathbb{R}^n$ satisfies polynomial identities, to appear.
77. P. M. Cohn, "Skew Field Constructions," London Math. Soc. Lecture Note series 27, Cambridge University Press, Cambridge/London/N.Y./Melbourne, 1977.
78. P. M. Cohn, The universal field of fractions of a semifir, *Proceedings London Math. Soc.*, to appear.